

GenCore version 5.1.4-p5-A578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: April 28, 2003, 12:00:02 ; Search time .259 Seconds

(without alignments)
6576.213 Million cell updates/sec

Title: US-10-087-013-1

Perfect score: 19455
Sequence: 1 atgggggtctctcgaataa.....atatatgatgaagtaattg 10628

Scoring table: BLOSUM62

Xgapop 10.0 ,	Xgapext 0.5
Ygapop 10.0 ,	Ygapext 0.5
Fgapop 6.0 ,	Fgapext 7.0
Delop 6.0 ,	Delext 7.0

Attached: 301932 segs, 80129803 residues

Total number of hits satisfying chosen parameters: 603864

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame_n2p.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US10087013/runat_28042003.102848.20384/app_query.fasta_1.10823
-DB=Published Applications_AA -OFMT=fastan -SUFFIX=n2p.rapp -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -STRAP=1 -END=1 -MATRIX=blomsu62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10087013.@CGN_1.1.119.@runat_28042003.102848.20384
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA:*

1:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2:	/cgn2_6/ptodata/1/pubpaa/PCF_NEW_PUB.pep:*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7:	/cgn2_6/ptodata/1/pubpaa/PCFUS_PUBCOMB.pep:*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12:	/cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	2972.5	15.3	2710	US-10-153-273-12
2	1032.5	5.3	700	US-10-153-273-10
3	665.5	3.4	1086	US-09-924-154-15
4	649.5	3.3	921	US-10-153-273-8

5	618	3.2	1435	9	US-10-153-273-4	Sequence 4, Appl1
6	607	3.1	1501	10 <td>US-09-924-154-17</td> <td>Sequence 17, Appl</td>	US-09-924-154-17	Sequence 17, Appl
7	600.5	3.1	1421	10 <td>US-09-924-154-13</td> <td>Sequence 13, Appl</td>	US-09-924-154-13	Sequence 13, Appl
8	597	3.1	1143	10 <td>US-09-924-154-14</td> <td>Sequence 14, Appl</td>	US-09-924-154-14	Sequence 14, Appl
9	431	2.2	749	9 <td>US-10-153-273-6</td> <td>Sequence 6, Appl1</td>	US-10-153-273-6	Sequence 6, Appl1
10	382.5	2.0	972	10 <td>US-09-924-154-16</td> <td>Sequence 16, Appl</td>	US-09-924-154-16	Sequence 16, Appl
11	365	1.9	411	9 <td>US-10-153-273-20</td> <td>Sequence 20, Appl</td>	US-10-153-273-20	Sequence 20, Appl
12	343.5	1.8	1115	9 <td>US-10-153-273-2</td> <td>Sequence 2, Appl1</td>	US-10-153-273-2	Sequence 2, Appl1
13	304.5	1.6	6281	10 <td>US-09-815-242-12996</td> <td>Sequence 12996, A</td>	US-09-815-242-12996	Sequence 12996, A
14	291	1.5	2125	10 <td>US-09-919-172-29</td> <td>Sequence 29, Appl</td>	US-09-919-172-29	Sequence 29, Appl
15	289	1.5	1639	9 <td>US-10-087-464-10</td> <td>Sequence 10, Appl</td>	US-10-087-464-10	Sequence 10, Appl
16	279	1.4	362	9 <td>US-10-153-273-18</td> <td>Sequence 18, Appl</td>	US-10-153-273-18	Sequence 18, Appl
17	279	1.4	2789	10 <td>US-09-801-574-57</td> <td>Sequence 57, Appl</td>	US-09-801-574-57	Sequence 57, Appl
18	276.5	1.4	2285	10 <td>US-09-932-183A-2</td> <td>Sequence 2, Appl1</td>	US-09-932-183A-2	Sequence 2, Appl1
19	276	1.4	411	9 <td>US-10-153-273-19</td> <td>Sequence 19, Appl</td>	US-10-153-273-19	Sequence 19, Appl
20	273.5	1.4	311	9 <td>US-10-153-273-21</td> <td>Sequence 21, Appl</td>	US-10-153-273-21	Sequence 21, Appl
21	270.5	1.4	1805	9 <td>US-09-820-843A-73</td> <td>Sequence 73, Appl</td>	US-09-820-843A-73	Sequence 73, Appl
22	249.5	1.3	2785	10 <td>US-09-801-574-8</td> <td>Sequence 8, Appl1</td>	US-09-801-574-8	Sequence 8, Appl1
23	243	1.2	2478	10 <td>US-09-815-242-5816</td> <td>Sequence 5816, Ap</td>	US-09-815-242-5816	Sequence 5816, Ap
24	243	1.2	2478	10 <td>US-09-815-242-12967</td> <td>Sequence 12967, A</td>	US-09-815-242-12967	Sequence 12967, A
25	232	1.2	3158	10 <td>US-09-815-242-12611</td> <td>Sequence 12611, A</td>	US-09-815-242-12611	Sequence 12611, A
26	228	1.2	2139	9 <td>US-10-023-219-4</td> <td>Sequence 4, Appl1</td>	US-10-023-219-4	Sequence 4, Appl1
27	228	1.2	2139	10 <td>US-09-727-384-6</td> <td>Sequence 6, Appl1</td>	US-09-727-384-6	Sequence 6, Appl1
28	226	1.2	2268	10 <td>US-09-815-242-5635</td> <td>Sequence 5635, Ap</td>	US-09-815-242-5635	Sequence 5635, Ap
29	226	1.2	2368	10 <td>US-09-815-242-12389</td> <td>Sequence 12389, A</td>	US-09-815-242-12389	Sequence 12389, A
30	222.5	1.1	665	9 <td>US-09-820-843A-107</td> <td>Sequence 107, Appl</td>	US-09-820-843A-107	Sequence 107, Appl
31	220	1.1	1331	9 <td>US-10-087-464-50</td> <td>Sequence 50, Appl</td>	US-10-087-464-50	Sequence 50, Appl
32	219.5	1.1	2025	10 <td>US-09-815-242-5703</td> <td>Sequence 5703, Ap</td>	US-09-815-242-5703	Sequence 5703, Ap
33	218.5	1.1	324	9 <td>US-10-153-273-17</td> <td>Sequence 17, Appl</td>	US-10-153-273-17	Sequence 17, Appl
34	217.5	1.1	5795	10 <td>US-09-815-242-12610</td> <td>Sequence 12610, A</td>	US-09-815-242-12610	Sequence 12610, A
35	214	1.1	807	9 <td>US-09-820-843A-108</td> <td>Sequence 108, App</td>	US-09-820-843A-108	Sequence 108, App
36	209	1.1	2434	10 <td>US-09-815-242-5835</td> <td>Sequence 5835, Ap</td>	US-09-815-242-5835	Sequence 5835, Ap
37	206	1.1	1400	10 <td>US-09-764-176-7</td> <td>Sequence 7, Appl1</td>	US-09-764-176-7	Sequence 7, Appl1
38	206	1.1	1404	10 <td>US-09-811-045A-1</td> <td>Sequence 1, Appl1</td>	US-09-811-045A-1	Sequence 1, Appl1
39	205	1.1	2835	9 <td>US-09-885-555-4</td> <td>Sequence 4, Appl1</td>	US-09-885-555-4	Sequence 4, Appl1
40	204.5	1.1	1183	9 <td>US-09-870-755-45</td> <td>Sequence 45, Appl</td>	US-09-870-755-45	Sequence 45, Appl
41	204	1.0	2437	10 <td>US-09-815-242-5834</td> <td>Sequence 5834, Ap</td>	US-09-815-242-5834	Sequence 5834, Ap
42	203	1.0	1719	9 <td>US-10-012-895-378</td> <td>Sequence 378, App</td>	US-10-012-895-378	Sequence 378, App
43	203	1.0	1719	9 <td>US-09-895-793-378</td> <td>Sequence 378, App</td>	US-09-895-793-378	Sequence 378, App
44	203	1.0	1719	9 <td>US-09-895-814-378</td> <td>Sequence 378, App</td>	US-09-895-814-378	Sequence 378, App
45	203	1.0	1719	10 <td>US-09-759-143-378</td> <td>Sequence 378, App</td>	US-09-759-143-378	Sequence 378, App

ALIGNMENTS

RESULT 1
US-10-153-273-12
Sequence 12, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Slim, Kim L.
Chiltnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Knodbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273

FILING DATE: 21-May-2002
 CLASSIFICATION: <Unknown>
 APPLICATION DATA:
 APPLICATION NUMBER: US/09/210,288
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Fuller, Michael
 REGISTRATION NUMBER: 36,516
 REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2710 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium falciparum
 SEQUENCE DESCRIPTION: SEQ ID NO: 12:
 US-10-153-273-12

Alignment Scores:
 Prod. No.: 1,9e-182 Length: 2710
 Score: 2972.50 Matches: 865
 Percent Similarity: 40.06% Conservative: 442
 Best Local Similarity: 26.51% Mismatches: 873
 Query Match: 15.28% Indels: 1083
 Gaps: 120

US-10-087-013-1 (1-10628) x US-10-153-273-12 (1-2710)

25 ATTATTAATGAGGATCAGATCATCATTAAGAGGAGATGCTTAAGACCCATATATA 84
 6 LeuAlaLysMetGlyProLysGluAlaIleGlyLysAspValIleGluAsp----- 22
 85 AAGAAAGTCACAAAGTCAGAAATGTTTGAACGTTTATGCCAAATTAAGACAT 144
 23 -----GluSerLalysHisMetPheAspArgIleGlyLysAspValIleAsp 38
 145 CCATCAAAA---TATGCAAAAGACATGTCGATTGTCGAAAGGGATTCGCAAGAA 201
 39 LysValLysGluGluAlaLysGluArgLysGlyLeuGlnIleArgLeuSerGluAla 58
 202 GAATTCGTGGTGTCTCTTACAGCCACTAATAAGCATATTTATTCATCAATCCA 261
 59 LysPheGluLysAsnGluSerAspProGlnThrProGluAsp-----Pro 73
 262 TGTATTTAGATCATTAAGAACATACATAATTACGATATGATGATGAATTCAGACAT 321
 74 -----CysAspLeuAspHisLysTyrHisThrAsnVal-----ThrThrAsnValIleAsn 90
 322 CTTGGCCATGCTAGAGAAACAAACCCGATTCATAGATGAAGAACTGATGATG 378
 91 ProCysAlaAspArgSerAspValAlaArgPheSerAspLeuIleGlyGlnCysThrHis 110
 379 AATTAATACGATTAATAAGAAATAAGATGATACCTTACCTGGGCGACACCTAGAGA 438
 111 AsnArgIleLysAspSerGlnGlnIleAspAsnLysGlyAlaCysAlaProIleArgAla 130
 439 CGACATATGCTGCTATAAAACCTTGGAAGCTCTAATGATTAATAATCCCAAAATATTCAT 498
 131 LeuHisValCysAspGlnAsnLeuGlnIleGluProIleLysIleThrAsnThrHis 150
 499 GATTATTTGGGAATGCTAGTACTAGTACGAAATAGCAAAATAGCAAGTAAATTCATTAAT 558
 151 AsnLeuLeuValAspValCysMetAlaIleLysPheGlnGlnIleThrGlnAsp 170
 559 CATCCA---CATAAAGGACT-----TCAGACGCTGTACTGCTCTTGA 600

171 TyrProLysTyrGlnAlaThrTyrGlyAspSerProserGlnIleCysThrMetLeuAla 190
 601 CGAAGTTTTCAGATATAGTATGATTTGTAAGAGAAATAGATATGTTAAACCAATGTC 660
 191 ArgSerPheAlaAspIleGlyLysPheValAlaArgGlyAspLeuTyrLeuGlyLysPro 210
 661 CATGAC-----AAAGTAGAAACGGGTCTCCGAGAGAGTTCGCAAAATA 705
 211 GlnGlnIleLysGlnArgGlnIleGlnLysAsnLeuLysThrIlePheGlyLysIle 230
 706 CATGAT-----GGAATGGAAGATGAAGTAAATAATGATTCATTCATGATGATCT 756
 231 TyrGluLysLeuAsnGlnLysAlaGlnAlaIleArgTyrGlyLysAsp-----ProGlu 246
 757 GGAATATTTAATTAATTAAGAGACATGTCGGAATGATGATTAAGTAAATTAAGTGGAA 816
 247 -----PhePheLysLeuArgGlnAspArgPheThrAlaAsnArgGluThrValIlePhe 264
 817 GGTATTAATGATGATGATCATATTAATTAATTCGATATTTATGCAATCGAAAGTAAATACA 876
 265 AlaIleThrCysAsnAla---TyrGlyAsnThrTyrPhe---HisAlaThrCysAsnArg 282
 877 CCATTAATTTCAATTCCTAATGCGGCATTAACACAGAAAGTTCCTACCAATTTAGAT 936
 283 GlyLysArgThrLysGlyTyrCysArgCysAsnAspAspGlnValProThrTyrPheAsp 302
 937 TATGTCCTCATATTTTACGTTGTTGTCGACAAATGGGAGAAAGATTTCCGCAAAAGA 996
 303 TyrValProGlnIleLysArgTyrPheGlnGluIleTyrAlaGlnAspPheCysArgLys 322
 997 AATTAATTTGTAAGAAAGTCAGAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1044
 323 AsnLysLysIleLysAspValLysArgAsnCysArgLysAspLysGlnLysAsp 342
 1045 CGCTATATTTGTGCTAATTAAGACATGATTTGACACACTATTTGCAAAAGATAT 1104
 343 Arg---TyrCysSerArgAsnGlnIleTyrAspCysGluLysThrLysArgAlaIleLys 361
 1105 TTCAATTTGGATTAATTAAGTACTGACTGCTGACATTAATGCAAAAGTTTGAAGTTGG 1164
 362 LeuArgTyrGlyLysGlnCysIleSerCysLeuTyrAlaCysAsnProIleValAspThr 381
 1165 TTGGGAATTCACACAGACATTTAAACAAACAAACAAACAAACAAACAAACAAACAAAC 1224
 382 IleAsnAsnGlnLysGlnIlePheAspLysGlnLysLysLysLysLysLysLysLys 401
 1225 TCATAT-----TATGCAACGATTAACAAATTTGCTCAAT 1257
 402 LysTyrGlnAsnGlnAlaSerGlyLysArgGlnLysArgAspAlaGlyThrThr 421
 1258 AATTAATTAATGTAAGAA---TATTAATTAACAAATTTTGAACAACTTAAGAAACCAATAT 1314
 422 ThrThrAsnTyrAspGlyTyrGlnLysLysPheTyrAspGlnLeuAsnLysSerGluTyr 441
 1315 GCAACTAATGACACTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1362
 442 ArgThrValAspLysPheLeuGlnLysLeuSerAsnGlnIleCysThrLysValLys 461
 1363 -----AAAGCGATTCACGAGAAAGATTAATTTACTTACTAACAGTCTGCT----- 1410
 462 AspGlnGlnGlyGlyThrIleAspPheLysAsnValAsnSerAspSerThrSerGlyAla 481
 1411 -----GATGACAAAGGATATTTATGCTTACGATTAATTTGCAAAATGTCGAAAGTGT 1455
 482 SerGlyThrAsnValGlnIleSerGlnIleThrPheTyrArgSerLysTyrCysGlnProCys 501
 1456 CCGCAGCTCGGGGTCGAATGATGATGATTAATTAATTAATTAATTAATTAATTAATTAAT 1512
 502 ProTyrCysGlyValLys-----LysValAsnAsnGlnLysLysSerSerAsnGln 517
 1513 CGTGACGCTGAATTAAT-----GAAGACTTAATTAACCTCATGAGGGGT 1554
 518 TrpGlnGlnLysAsnAsnGlnLysCysLysSerGlyLysLeuTyrGlnProLysProAsp 537

QY 1555 GTGAACCTACTAATATCACTGCTCTTATAGTGGATGAACAGCTGATATTACAA 1614
DB 538 LysGluGlyThrThrIleThrIleLeuLysSerGlyLysGlyHisAspSeriLeuGlu 557
QY 1615 AATATGAATAATTTTGTAC-----ACCTCA 1641
DB 558 LysLeuAsnLysPheCysAspGluLysAsnGlyAspThrIleAsnSerGlyLysGly 577
QY 1642 ACTAATACAAAGATTAATAAT-----CAAAATGGAGATG 1680
DB 578 ThrGlyGlySerGlyGlyLysAsnSerGlyArgGlnLeuLeuThrGlnGluTrpLysCys 597
QY 1681 TATTAATAGAGTAAATATTAATAGATGAATCACTGAACAAATCTGAAATCAATAT 1740
DB 598 ---TyrLysGlyGluAspValValLysValGlyHisAspGluAspGlnGluAspTyr 616
QY 1741 GATTAACCTAG-----ATAATA----- 1758
DB 617 GluAsnValLysAsnAlaGlyGlyLeuCysIleLeuLysAsnGlnLysLysAsnGly 636
QY 1759 -----TCATTTCATTAATTT 1773
DB 637 GluGlyGlyAsnThrSerGlyLysGluProAspGluIleGlnLysThrPheAsnProPhe 656
QY 1774 TTGGAATTTGGGTACATATTATTAAGGAGTACTATTAGTGAATGACAACTTAA 1833
DB 657 PheTyrTyrTrpValAlaHisMetLeuLysAspSerIleHisTrpLysLysLeuGln 676
QY 1834 ACTGTAT-----ATAATCAACACGCACTGT-----ATGATGAATGAACAAATTC 1887
DB 677 ArgCysLeuGlnAsnLysAsnArgIleLysCysGlyAsnAsnLysCysAsnAspCys 696
QY 1888 TTAATGTTTTCAGAGATGGTTAAACAAAGAGAAATGAATGATTAAGAAACATG 1947
DB 697 GluCysPheLysArgTrpIleThrGlnLysLysAspGluTrpLysIleValGlnHis 716
QY 1948 TTCACAAAAGAAAGATATACG----- 1971
DB 717 Phe---LysThrGlnAsnLysLysGlyArgGlySerAspAsnThrAlaGluLeuIle 735
QY 1972 -----CAATCGTATATAGTATATTAATTAATCTTTTGAAGGTATTTT----- 2016
DB 736 ProPheAspHisAspTyrValLeuGlnTyrAsnLeuGlnGluLysLeuLysGlyAsp 755
QY 2017 -----TTTAAAGTTATGATTAACCTGACAAAGATGAAGCAAAATGCG 2058
DB 756 SerGluAspAlaSerGluGlnLysSerGluAsnSerLeuAsnAlaGluGluAla----- 773
DB 2059 AAAGAACTTATGAAATATTAATAAAGAAATAATGAGTTTCCAAATTTGAAATAT 2118
DB 774 -----GluGluLeuLysHisLeuArgGluIleIleGlnSerGluAspAsnAsp 789
QY 2119 AGGAC-----TATTAGAGATGCATACCACTTGT 2154
DB 790 GlnGluAlaSerValGlyGlyValThrGlnGlnLysAsnIleLysLysLeuLeu 809
QY 2155 GATCACTTAAGAAAGATCCACAGATATG----- 2184
DB 810 AsnTyrGluLysAspGluLysLeuLysCysLeuGlnIleHisGlnAspGlnGluGln 829
QY 2185 ---AAAGACAAATTAATCAAGACAGATGTGAACATCCCATATATGCAACAAACCG 2241
DB 830 LysGluLysGlyAspLysAsnGlnCysIleGlnGlnGluLysAsnPheArgTyrAsnPro 849
QY 2242 TGTGTTAAACCTGCGAGCAGCAAA---CCCACTAAATAATTAAGAAATAGACAA 2298
DB 850 CysSerGlyGluSerGlyLysAsnLysArgTyrProValLeuAlaAsnLysValAlaTyrGln 869
QY 2299 TACTTTAAAGAGTCATACGAGAGACAGCAAGAACTGCGCTTCTCATTAATTAAGAA 2358
DB 870 MetHisIleLysAlaLysThrGlnLeuAlaSerArgAlaGlyArgSerAlaLeuArgGly 889

QY 2359 AAGCACACAGAGCTATATTAACGTGGGGTAGAGAAAGAAAGACTTCAAGCAATTA 2418
DB 890 AspIleSerLeuAlaGlnPheLysAsnGlyArgAsnGlySerThrLeuLysGlnIle 909
QY 2419 TGTGAATATATGAATAAAGCTTAATCGTAATCTGCTTTTCAATGACAGATGAT 2478
DB 910 CysLysIleAsnGluAsnTyrSerAsnAspSerArgGlyAsnSerGlyGlyProCysThr 929
QY 2479 GGCAAAGCAGCAGCTGATGATTAACAACAAAGATTGCTGATGAAGATGAAGGAAAGT 2538
DB 930 GlyLys---AspLysAsnHisGlyGlyValArgMetArgIleGlyThrGluTrpSerAsn 948
QY 2539 GATCCGACACATCGCTAAAGATCACAGAAATGTTATTAATGCTCTTACAGACAT 2598
DB 949 IleGluGlyLysLysGlnThrSerTyrLysAsnValPheLeuProPheArgGluHis 968
QY 2599 ATATGTATACCAATTTGGAACTTTACAAAGAGATGATCCACACTTAATGGTAATAT 2658
DB 969 MetCysThrSerAsnLeuGluAsnLeuAspVal-----GlySerVal 982
QY 2659 GTT---GATGATTTAGTTAATAATCTTTTGGGGATGCTCTTATCAGCAAAATAT 2715
DB 983 ThrLysAsnAspLysAlaSerHisSerLeuLeuGlyAspValGlnLeuAlaIleLysThr 1002
QY 2716 GAGCAACACAGTATATACAGATGTATTAAGAAAGATTAACSTAAAGGCCCAAGAA 2775
DB 1003 AspAlaAlaGluIleLysArgTyrLysAspGlnAsnAsnIle-----Gln 1018
QY 2776 GTAACTGACCCA-----AAACACACACAACTATGCTGATACGTATACGTACGT 2826
DB 1019 LeuThrAspProIleGlnGlnLysAspGlnGluAlaMetCysArgAlaValArgTyrSer 1038
QY 2827 TTTGCAATATAGATATATTAATTCGAGAAAGATCTGCGAAAGAAACGCTGACATG 2886
DB 1039 PheAlaAspPheLysAspIleIleArgGlyArgAspMetTrpAspLysSerSer 1058
QY 2887 GTAAAGCTGCAGACATTTGGAAGCTTTTGGTATATATACATTAAGTCACTCAAGC 2946
DB 1059 ThrAspMetGlnThrArgLeuIleThrValPheLysAsnIleLysGlnLysHisAspGly 1078
QY 2947 ---AAAGAAATGATTAATTAATGATGAT-----GCCCCAAATTTTAAATTTG 2994
DB 1079 IleLysAspAsnProLysTyrThrGlyAspGlnSerLysLysProAlaTyrLysLeu 1098
QY 2995 AGCGAAATGTTGCGAAGCTAATAGACCCAAAGTATGGCAACATGAATGCTGATATA 3054
DB 1099 ArgAlaAspTrpTrpGlnAlaAsnArgHisGlnValTrpArgAlaMetLysCys----- 1116
QY 3055 AATATTTGAAGATTAATGGGACACAACTCAACACAAAGTATATTCGATATAGT 3114
DB 1117 -----AlaThrLysGlyIleIleLys----- 1123
QY 3115 GATCATACACATGATGATATATCCACAAATTAAGATGAGTACGGAATGGCA 3174
DB 1124 ProGlyMetProValAspAspTyrIleProGlnArgLeuArgTrpMetGlnTrpAla 1143
QY 3175 GAATGCTACTCAAGGTCGAAAGAAAGAGTATGATTAAGTGAAGAGATGAAGAG 3234
DB 1144 GluTrpTyrCysLysAlaGlnSerGlnGluTyrAspLysLeuLysLysIleCysAlaAsp 1163
QY 3235 TGTAAAGATTAAGATATGCTCAAGGCTGACAAAGACAGTGTACAGGTTGTACAG 3294
DB 1164 CysMetSerLysGlyAspGlyLys---CysThr---GlnLysAspValAspCysGlyLys 1181
QY 3295 TGCACAGAGAGCTGATATGAATATGAATATGAATATGAATATGAAGAAAGATGAAT 3354
DB 1182 CysLysAlaAlaLysAspLysTyrLysGlnGluIleGlnLysTrpAsnGlnGlnTrpArg 1201
QY 3355 ATATATACAGATTAATACAAAGATTTACAGAAACAGCAACAAATGCTGTACTAATATG 3414
DB 1202 LysIleSerAspLysTyrAsnLeuLeuTyrLeuGlnAlaLysThrThrSerThrAsnPro 1221
QY 3415 GGTATTGAAGCTTCAGTACTGCCAAATAATCATATAGAC-----AGGAATGTTAT 3465

QY 5563 ACTGCTATCAGTTTCTTGATGCTTTGGCGAATGGGCTGAAGATTTTTGCACAACTAAA 5622
Db 1701 ----- 1701
QY 5623 GAAAAGAAATTGGAGAAATTGGTAGGGCGGTGAATGATTACTTGTGCTGATATGAA 5682
Db 1701 ----- 1701
QY 5683 GATTAAGAAAGAAATGTACAGATGCGGTACACAAATATATAAAAATTATTAAGTAGTG 5742
Db 1702 ----- 1704
QY 5743 AAACCCAGATGAAAAAACAATCAAAAATATGTTGAGAAATTAAGACAAATATATTC 5802
Db 1704 ----- 1704
QY 5803 GAGCATCTGTGGCAAAAGATGACAGAGCGTCGCAATATTATAGACAAACAAATTA 5862
Db 1704 ----- 1704
QY 5863 AAAATTGTGAAAAATTAAGTGAGATTGTGAATATATAGTATGAAGATGTGTCCACA 5922
Db 1705 ----- 1706
QY 5923 CAGCGATTACTGATGATATAGTCAAAATATGCCGCATCATATAGACATGACCCAAA 5982
Db 1706 ----- 1706
QY 5983 GAAGTTGAAGAAAGTAAATTGTCAAGTGCACAGAGTCCACAGCTGTAGCAAGGAA 6042
Db 1707 ----- 1717
QY 6043 ACACGCTACACAGGGGTACTGATATCAAAAACGAGCGCATGAAAAAGAGCGAAA 6102
Db 1717 ----- 1717
QY 6103 ACAGCGCGGCTACAAACAGCGGAAAAAGTGAATCTAACACAGAAATGCGAGCA 6162
Db 1718 ---ProProProProValInProGln----- 1725
QY 6163 CAACACAGAACCCGACAGACACACACAAACAGAAACATCAACAGCAACACA 6222
Db 1726 ----- 1730
QY 6223 ACAGAACTGAGCTGGGCACAATGTAAAGCCATTCTTGAATTAACACAGATAGCAG 6282
Db 1731 ValThrValAspValCysSerIleValIysThrLeuPheLys-----AspThrAsn 1747
QY 6283 GGTGAATAGAGGCTGTAATCCAAAGATGACATAT-----CCTAAATGGGCT 6336
Db 1748 AsnPheSerAspAlaCysIleLeuLys-----TyrGlyLysThrAlaProSerSerIlePlys 1766
QY 6337 TGTATT-----GTAGGTAAAGCTTAAAGAAAAT 6363
Db 1767 CysIleProSerAspThrLysSerGlyAlaGlyAlaThrThreGlyLysSerGlySerAsp 1786
QY 6364 GAAATATGATATGATGCTCTCTAGGAGAAAAAATATGATTAATATATATATATAT 6423
Db 1787 SerGlySerIleCysIleProArgArgArgLeuTyrValGlyLysLeuGln--- 1805
QY 6424 TTAATATGAACCT-----GAAAATTAAGCTGAC 6453
Db 1806 ---GluThrAlaThrAlaLeuProGlnGlyGlyAlaAlaProSerHisSerAlaGlyAla 1824
QY 6454 AATGATATAAAGAGCTTTTATTAATGTGACAGATAGAACTCAATTTTGTGCTTA 6513
Db 1825 AspAspLeuArgAsnAlaPheIleGlnSerAlaAlaIleGluThrPhePheLeuTyrPasp 1844
QY 6514 AATATATATATGAA---AATCTGACAGCAAAATGAA----- 6549
Db 1845 ArgTyrLysGluGluLeuLysLysProGlnGlyAspGlySerGlnGlnAlaLeuSerGlnLeu 1864

QY 6550 -----TTGCAAAATGAGACA 6564
Db 1865 ThrSerThrTyrSerAspAspGluIleAspProProAspLysLeuGlnSerGlnLys 1884
QY 6565 ATTCAGATGAATTTTAAAGAAATATGATATATACATATATGATATATATATATATAT 6624
Db 1885 IleProProAspPheLeuArgLeuMetPheTyrThrLeuGlyAspTyrArgSpIleLeu 1904
QY 6625 TTGGAACTGATATTTCTAATGATTAATAAATTAATTAACGTAAACAAATAGCTAACAC 6684
Db 1905 ValHisGlyGlyAsnThrSerAspSerGlyAsnThrAsnGlySerAsnAsnAsnIle 1924
QY 6685 ATTCGCAATGAAAAATTAAGAAAAACAGATTAATAAAGAAATGACAAATTAACCTAA 6744
Db 1925 ValLeuGluAlaSerGlyAsnLysGluAspMetGlnLysIleGlnGluLysIleGln 1944
QY 6745 ATA----- 6747
Db 1945 IleLeuProLysAsnGlyGlyThrProLeuValProLysSerSerAlaGlnThrProAsp 1964
QY 6748 ---TTTGGAGAAAAATTAATAAATTTATTTGGAGAAAGATATATATGATTAACCTTA 6804
Db 1965 LysTrpTrpAsnGlnHisAlaGlnSerIleTrpLysGlyMetIleCysAlaLeuThrTyr 1984
QY 6805 -----CATCTCACAGACGAAACGAAACGAAACGAAACGAAACGAAACGAAAC 6840
Db 1985 ThrGluLysAsnProAspThrSerAlaArgGlyAspGluAsnLysIleGluLysAspAsp 2004
QY 6841 GATTAATTAC----- 6849
Db 2005 GluValTyrGluLysPhePheGlySerThrAlaAspLysHisGlyThrAlaSerThrPro 2024
QY 6850 -----CAGTCAAT---GACATGACCAAACTG--- 6873
Db 2025 ThrGlyThrTyrLysThrGlnTyrAspTyrGluLysValLysLeuLysAspThrSerGly 2044
QY 6874 -----ACGCTCCCTGGAAGAGCTTTGTAATAAAGCCG 6906
Db 2045 AlaLysThrProSerAlaSerSerAspThrProLeuLeuSerAspPheAlaLeuArgPro 2064
QY 6907 CAATTTTGAGATGCTTCAACAGATGGCAGAAAGAAATTTGTATATAGAGAAAGAACAG 6966
Db 2065 ProTyrPheArgTyrLeuGluGluTyrPglGlnAsnPheCysLysLysArgLysHisLys 2084
QY 6967 TTGTTAAATTTGAGGCGGCTGTAAAGAAATGAG----- 7002
Db 2085 LeuAlaGlnIleLysHisGlnCysLysValGluGluAsnGlyGlyLysSerArgArgGly 2104
QY 7003 -----TGTAAAT-----GCTACTAAT 7017
Db 2105 GlyIleThrArgGlnTyrSerGlyAspGlyGluAlaCysAsnGluMetLeuProLysAsn 2124
QY 7018 GACGCTAGACACAGAA-----TGTGACAGAGCGTGTATACATATCA 7062
Db 2125 AspGlyThrValProAspLeuGluLysProSerCysAlaLysProCysSerSerTyrArg 2144
QY 7063 AATTTATTAAGAGTGAAGAACTGAATATGAAGACAAAGAAAGAAAGTTCAAAAGAT 7122
Db 2145 LysTrpIleGlnSerLysGlyGluPheGlnGlyGlnGluLysAlaTyrGlnGlnGln 2164
QY 7123 AAAGAT----- 7128
Db 2165 LysAspLysCysValAsnGlySerAsnLysHisAspAsnGlyPheCysGluThrLeuThr 2184
QY 7129 ---GGCAAAAGATATAGATATAT-----CCTTCTACTGAAAGAGACATA 7170
Db 2185 ThrSerSerLysAlaLysAspPheLeuLysThrLeuGlyProCysLysProAsnAsnVal 2204
QY 7171 GACAGGCAACATGTCTCATGCAATATTAACATGAAATTAAGAAATATATGTGCAAT 7230
Db 2205 GluGlyLysThrIlePheAspAspAspLysThrPheLys-----HisThr 2219
QY 7231 AAGGATTTCTTCTGATGCAAAAACCTTCTTCAACAACCTTCAACAACAAACACACACATCA 7290

```

Db 2220 LysAspCysAspProCysLeuLysPheSerValAsnCysLysLys----- 2234
QY 7291 CATCATCCGATGCTAATGATATGCGCAATACCGCTGGATTATGTTCTCGAAGAAATTAAAC 7350
Db 2235 -----Asp 2235
QY 7351 AAGTGTAGCTGCTGCAACTTTCAAAAAAGGAGCTATGATTCATACAAAAAATTACT 7410
Db 2236 GlucylAsp-----AsnSerLysGlyThrAspCysAlaGlnLys----- 2248
QY 7411 GAACCTAAATATACCTATGATTTGTGTAGAAACAGACATATATTTATCTTAAGAACCA 7470
Db 2249 -----AsnSerLysAsp-----AlaThrAspIle 2256
QY 7471 GAAATATATGATATATCTTCAAGGAAAAAATTATACCTATTAGTGTACAAAGAA 7530
Db 2257 GluAsnGlyValAspSerThrVal-----LeuGlnMetAlaGlyAlaSer 2270
QY 7531 AAGAAAGTAAATATAGTGTGACTAATAATATACCTCGATCCTTAAGAAACCTTATGCA 7590
Db 2271 AlaAspSerLysSerGlyPheAsnGlyAsp----- 2280
QY 7591 CCTGATATATATATAGGAAAGAAACCTTGTGAAAT----- 7629
Db 2281 -----GlyLeuGluAsnAlaCysArgIlyAlaGlyIlePheGluGlyIle 2295
QY 7630 AGAGAAAGAAATCGTTTAAAGTAGATATGAAATGGAATGTTCACAAAATTCAAAATGC 7689
Db 2296 ArgLysAsp-----GluThrLysCys-----ArgAsnValCysGly 2307
QY 7690 TATCAGGAGAAAAAAGAGTAGTGTACCTCCCAAGAAAGAACATATGTGCTTAAGAAAT 7749
Db 2308 TyrVal-----ValCysLysPro----- 2313
QY 7750 TTGATGCAATTTAAATGTGAAGACTTAAAGATAGTATATATCTCTAAAAAATGTTGCT 7809
Db 2313 ----- 2313
QY 7810 CGAATCGACGAAATGAGAGAAATAGACATATATAAACTTCACTCAGACAGACGGGTGC 7869
Db 2314 -----GluAsnValAsnGlyGlu----- 2319
QY 7870 GCAATGAATCCAAATATGATGATCTATGAAATATAGTTCGCTGATCGGTGACATAGTT 7929
Db 2320 -----Ala 2320
QY 7930 AGAGGAACAGATATGTTAGCAATGTGTGTTACTTACCTCCCGTAGAAATAAATTATAT 7989
Db 2321 LysGlyLysHisIleIleGlnIleArgAlaLeuValLys-----ArgTyrVal 2336
QY 7990 AAGCTTTTGAATACATATATGAAATATGAGAAATATAAAGTAGAGAAATTAATAC 8049
Db 2337 GluTyrPhePheGluAspTyrAsnLysIleLysHisLysIleSerHisArgIleLysAsn 2356
QY 8050 AACGAT-----GTCAAAACGTTTCTGCTCTGCTGGGAGTGAATAGAAAA 8097
Db 2357 GlyGluIleSerProCysIleLysAsnCysValGluLysTyrValAspGlnLysArgLys 2376
QY 8098 GATATTTGAAAGACATGACGTGCAAGACACAGACAGATGCAAACTTTTACAAAAAGA 8157
Db 2377 Glu-----TyrLysGluIleThrGluArgPheLysAsp----- 2387
QY 8158 AGAATGATGATTTGAACCGCATTAATTCANAGATAGTGAGACATAAGACGAT 8217
Db 2388 -----GlnTyrLysAsnAspAspAspAsp 2396
QY 8218 CCACCTGTGATGATATATATACCTCAACGGTTTCGATGATGACGTACAGTGTGAATAT 8277
Db 2397 AspAsnValAlaSerPheLeuGluThrLeuIleProGlnIleThrAspAlaAsn----- 2414
QY 8278 TATTGTAAACACCTGATGAGAAATTTGAAAAAATTAATCATGATCATCAGTATAA 8337

```

```

Db 2415 ---AlaLysAsnLysValIleLysLeuSerLysPheGlyAsnSerCysGly---CysSer 2432
QY 8338 ACATCGACAGATGCAAGATGATTATGATGAAAAATAGTGTGAACAGCTGTAAACGAGA 8397
Db 2433 AlaSer-----AlaAsnGluGlnAsnLysAsn----- 2441
QY 8398 TGTCAAAATATATAAATTTCTTCTTAAATGCAAAATCTCATTCGATATACAAATCAAT 8457
Db 2442 ---GlyGluTyrLysAspAlaIle-----AspCysMetLeuLys 2453
QY 8458 AATATCAAAAGATTTGTATGACACACCAATATATACMAAAATCTCTACTATGATCATGTT 8517
Db 2454 LysLeuLysAsp-----LysIleGlyGluCysGluLysLys 2465
QY 8518 CAAATTTTGTCAAAAGTGTGAACCTTTAAAGTAGAATGTTCTGTGGAGAGCTTTCT 8577
Db 2466 HisHis-----GlnThrSerAspThrGluCysSerAspThrProGlnPro 2480
QY 8578 GAATATCTTCATGAACAGTAACTGTTGAATTTATTAATTAAGAAATGATGTTCT 8637
Db 2481 GlnThrLeuGluAsp-----GluThrLeuAspAspIleGluThrGluGluAlaLys 2498
QY 8638 TCAATATACGACAAATATGCTTTCAGAAACACCAAAAGTTATTAAGAA----- 8688
Db 2499 LysAsnMetMetProLysIleCysGluAsnValLeuLysThrAlaGlnGlnGluAspGlu 2518
QY 8689 ---GCTTGCAGTTGTACACTCTTCAAGATTCATGATGATTAATGTTCTCTACCGATCA 8745
Db 2519 GlyLysCys-----ValProAlaGluAsnSerGluIleProAlaAlaThrAspSer 2535
QY 8746 AACAAAGT 8754
Db 2536 GlyLysGlu 2538

```

RESULT 2
 US-10-153-273-10
 Sequence 10, Application US/10153273
 Patent No. US20020169305A1
 GENERAL INFORMATION:

APPLICANT: Shim, Kim L.
 Chitnuls, Chetan
 Miller, Louis H.
 Peterson, David S.
 Su, Xin-zhaun
 Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:

ADDRESSSEE: Knobbe Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/153,273
 FILING DATE: 21-May-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/210,288
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:

NAME: Fuller, Michael
 REGISTRATION NUMBER: 36,516
 REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 700 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium falciparum
 SEQUENCE DESCRIPTION: SEQ ID NO: 10:
 US-10-153-273-10

Alignment Scores:
 Pred. No.: 3.5e-58 Length: 700
 Score: 1032.50 Matches: 247
 Percent Similarity: 50.288 Conservative: 110
 Best Local Similarity: 34.798 Mismatches: 250
 Query Match: 5.318 Indels: 103
 Gaps: 22

US-10-087-013-1 (1-10628) x US-10-153-273-10 (1-700)

```

OY 406 AATGATGCTATAGCCCTGCGCCACCTAGACAGCAGCATATGTCATATAAACTTGGAA 465
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5 AsnLysValGlyAlaCysAlaProTyrArgLeuHisLeuLysCysAspTyrAsnLeuGlu 24
OY 466 GCTCTAATGATATAATCCCAAAATATTCATGATTTATGGGAATGTAAGTACTTACA 525
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25 SerLeuAsp-----ThrThrSerThrThrHisLysLeuLeuGluValCysMetAla 42
OY 526 GCATAATACGAGAGTGAATCATATTTGTAATTAATCAT-----CCACATAAGAGCACT--- 576
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 43 AlaLysTyrGluLysAsnSerIle-----AsnThrHisTyrThrGlnHisGlnArgThrAsn 61
OY 577 -----TCAGACGCTTGACTGCTGCTGTCGACGAAAGTTTCCAGATATAGTGAT 624
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 62 GluAspSerAlaSerGlnLeuLysThrValLeuAlaArgSerPheAlaAspIleGlyAsp 81
OY 625 ATTGTAGAGAGTAATAGATATGTT-----AAACCAATATGTCATGAC 666
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 82 IleValArgGlyLysAspLeuTyrLeuGlyTyrAspAsnLysGluLysGlnArgLys 101
OY 667 AAAGTAGAAAAGGGGTCTCCGAGAGGTTTCAAGAAAATACATGATGAAATGAAATGAA 726
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 102 LysLeuGlnGluLysLeuLysAspIlePheLysGlyIleHis-----LysAspVal 118
OY 727 GTAAAAAATGAT-----TACAATCCCTGATGGA---TCTGGAATTTATAT 768
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 119 MetLysThrAsnGlyAlaGlnGluArgTyrIleAspAspAlaLysGlyLysAspPhe 138
OY 769 AAATTAAGAGAGCATGCGTAATGTAGTAAGTAATTAAGTATGGAAGCTATTAACATG 828
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 139 GlnLeuArgGluAspTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrp 158
OY 829 GATGCATCATATTAATCTGGATATTTTATGCAATCAGAAAGTAATACACCATATTTTGA 888
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 159 HisAlaProLysGluAlaAsnTyrPheIleLysThrAlaCysAsnValGlyLysGlyThr 178
OY 889 AATCCCAATATGCGGCATATAACAGAAAGTTCCCTACCAATTTATATATGCTCCCTCAA 948
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 179 AsnGlyGlnCysHisLysIleGlyLysAspValProThrPheAspTyrValProGln 198
OY 949 TATTTCGTTGTTGTCAGCAATGCGAGAGAGTTTGGCGAAAAGTAATATTAATG 1008
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 199 TyrLeuArgTrpPheGlnGluTrpAlaGluAspPheCysArgLysLysLysLysLysLys 218
OY 1009 AAAAAGCTAAGAGACTCTGCTGTAATGACAAAGAAAGCTTATATTTGATGTCATAATGA 1068
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 219 GluAsnLeuGlnLysGlnCysArgAspTyrGluGlnAsnLeuTyrCysSerGlyAsnGly 238
OY 1069 CATGATTTGACGACAACTATTTGGAAAAAAGTATTTTGCATTGGATATAATAGTCTACT 1128
  
```

```

DB 239 TyrAspCysThrLysThrIleTyrLysGlyLysLeuValIleGlyGlnHisCysThr 258
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1129 GACTGTCACATTAATGCAAAAGTTTGTGAAGTTTGGTATGGGAATACACAGACATTT 1188
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 259 AsnCysSerValItrpCysArgMetTyrGlnThrTrpIleAspAsnGlnLysGluPhe 278
OY 1189 AAAAACAACAAAAGAAAATATGCAAAAGAAATG----- 1221
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 279 LeuLysGlnLysArgLysTyrGlnThrGlnIleSerGlyGlyLysSerGlyLysSerPro 298
OY 1222 -----CATCATATTTTTCGACAGATACAAATTTGCAAT 1257
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 299 LysArgThrLysArgAlaAlaArgSerSerSerSerAspAspAsn----- 314
OY 1258 AATATTAATAGTCATATTAATTAACAAATTTTGAAGAACTTATGAGCAACGAAATATGCA 1317
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 315 -----GlyTyrGlnSerLysPheTyrLysLeuLysGluValGlyTyrGln 330
OY 1318 ACTAATGACACTTTTAAATTTACTTAATGAGAAAGTATTTGTAAGCA----- 1368
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 331 AspValAspLysPheLeuLysIleLeuAsnLysGluGlyIleCysGlnLysGlnProGln 350
OY 1369 ---GGATTACACAGAGAAAAGCATATTACTTTACTTAACAGTCTCATGCAAAAGGAGATA 1425
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 351 ValGlyAsnGlnLysAlaAspAsnValAspPheThrAsnGluLysTyrValLys---Thr 369
OY 1426 TTTTATCGTTGACATATATGCGCAAGTGTGCCGACTGGCGGGGCAATGATGATATA 1485
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 370 PheSerArgThrGluIleCysGlnProCysProTyrCysGlnLeuLysGlyLysPro 389
OY 1486 AATACACACACAAATACATATATGATGCTGAACGTGTAATTAAGTAAGCACTATAACCT 1545
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 390 ProTyrValLysGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 409
OY 1546 CCATGGGGGTGACGCTACTAATATCATGCTGCTTTATATGTCGTAATGCAAGGTGAT 1605
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 410 Lys-----AsnIleThrAspIleProValLeuTyrProAspLysSerGlnGlnAsn 426
OY 1606 ATTACACAAAATTAATTAATTTTGTACACAGCTCAACTAATTAACAGTATAAATAATAT 1665
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 427 IleLeuLysLysTyrLysAsnPheCysGluLysGlyAlaPro---GlyGlyGlnIle 445
OY 1666 CAAAATGCGAATGCTATTATTAAGAT-----CAAAATTAATTAAGATG 1710
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 446 LysLysTrpGlnCysTyrTyrAspGlnHisArgProSerSerLysAsnAsnAsnAsn 465
OY 1711 AAACGTGGAACAAATCTGAATCATATGATGATATCCATAGATATATCATTTTCATAT 1770
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 466 ValGluGlyThrTrpAspLysPheThrGlnGlyLysGlnThrValLysSerTyrAsnVal 485
OY 1771 TTTTGAATTTATGCGTTACATATTTATTAAGGATACATATTAAGGATGCAAACTT 1830
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 486 PhePheTrpAspTrpValHisAspMetLeuHisAspSerValGlnTrpLysThrGluLeu 505
OY 1831 AAAACTGTATATAATATACAAACACG-----CATGTATGATGAATGTAAC 1878
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 506 SerLysCysIleLeuAsnAsnAsnThrAsnGlyAsnThrCysArgAsnAsnAsnLysCys 525
OY 1879 AGAAATTCCTTATGTTTTCAGACAGATGGTTAAACAAAAGAAAGAAAGAAATGATATA 1938
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 526 ThrAspCysGlyCysPheGlnLysTrpValGluLysGlnGlnGlnGlnGlnGlnGln 545
OY 1939 AAGAACTGTACAAAAAAGAAATATACAGCAATCGATATATGTAATATATATAT 1998
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 546 LysAspHisPheGlyLysGlnThrAspIleValGlnGln-----LysGly 560
OY 1999 CTTTGAAGGATATTTTAAAGTATGATGATGATGATGATGATGATGATGATGATGATG 2058
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 561 LeuIleValPheSerProTyrGlyValLeuAspLeuVal-----LeuLysGly 576
OY 2059 AAAGACTTATGCAAAATATTAATAAAGAAAGAAAGAAAGTATTTCCATTTGGAAAAATAT 2118
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

Alignment Scores:	
Pred. No.:	1,32e-34
Score:	665.50
Percent Similarity:	33.55%
Best local Similarity:	21.02%
Query Match:	3.42%
Length:	1086
Matches:	312
Conservative:	186
Mismatches:	431
Indels:	555

[illegible]


```
Db 360 ----- 360
QY 7573 CTAAGAAACCTTATGCACCTGATAATATAGAGAGAAACCCTTGAAATAGA 7632
Db 360 ----- 360
QY 7633 GAAGAAATCGTTTAAAGTAGATTGAATGGAATGTTACAAAATTCAGAGTCTAT 7692
Db 361 -----Serlysalyslystrpsincyslyserinsalys---Val 375
QY 7693 CAGAGAAAAAAGAGTAGTGTACTCCAGAGAGAGAAACATATGCTTAAGAAATTA 7752
Db 376 ThlypProgluylalysglYpProProlArglnInleucysleuylYtylle 395
QY 7753 GATGAATTTAA-----ATTGAAGACCTTAAGAGTAGTAATATCTCTAAAAATG 7803
Db 396 PheleuIlleArgspcllysanglulglyLeuIlyAspHis----- 409
QY 7804 GTTCGTGAACTGCAGAAATGAGAAATAGACATATATAAAAACCTCACTCAGAGAAC 7863
410 IleaInlyshlaIlaInlyrGuIlaIleuHlsleuIyglYstrGuIlaInlaIgly 429
QY 7864 GGGTGGCAATGAATCAATATGTGATCAATATATAGTTCGTCGTGAC 7923
430 Gly-----AspIlyIleCysAsnAlaIleleuclYserTyrlaAspIleGlyAsp 446
QY 7924 ATAGTAGAGAACAGATATGTAGCAATGTGGTTACTTACCTCCGTAAGAAATP--- 7980
447 IleValArgglYleuAspValITrArg-----AspIleAsn 458
QY 7981 -----AATATATATAGGTTTTTGAATACATATATGAAATGAGAAATTAATA 8034
459 ThrAsnIlyserSerGlyIlyPheGlnIlyIlePheMetGlyGlyIlyAsnSerArgly 478
QY 8035 GGTAGAAATTAATAACAGATGTACAAACGTTCTCTCTGCGTGGAGTGTATAGA 8094
479 LysGlnInAspAsnAnglu-----ArgAsnIlystrpIrgIlyGlnArg 494
QY 8095 AAGATATTTTGAAGCAATG-----ACGTCAAGACACCA 8130
495 AsnleuIlleTrpserSerMetValIlyshIleProlYsglyIlystrCysIys----- 512
QY 8131 GAAGATGCAAACTTTTGAAGAAAGAGATGGATGGATTGACCATTAACATTATA 8190
513 -----ArgHisAsnAsnIleGlyIly----- 519
8191 CAAGATAGCTGACATAAGACATCCACCTGTGATATATATCTCAACGTTT 8250
520 -----IleProlInPheLeu 524
QY 8251 CGATGATGACATGATGCTGATATATTGTAAGACACTGATGGAATGGAAAA 8310
525 ArgTrpIleuIyglYtrpIrgIlyAspGluPheCysGluIleuMetGlyTrpGluValIlyGln 544
QY 8311 TTTAAAAAATCATGATGATCACTGTAAACATCTGACAGATGCAAGATGATTGATGA 8370
545 LeuGluIlyIleCysGluAsnIlyAsnCysSerGlyIly----- 557
QY 8371 AATAAGGTGAACGTATAAAGAGATGTCAGAAATATAAATTTTGTCTTAATG 8430
558 -----LysCysIlyAsnAlaCysSerSerTyrlYstrIleYsgluArg 573
QY 8431 AATCTCTATTCATATACATAAATTAATACAAAGAAATGTATGCAACCAATATAT 8490
574 LysAsnIlyrInserInserIlyIlyPheAspSer-----AspIlyIlyIleAsn 591
QY 8491 ACAAAAATCTCTACTATGATCATGTTCAAAATTTGTACAAAAGTGAACCTTTTAA 8550
592 LysIlyAsnAsnIleuTyraIn----- 598
QY 8551 AGGAATGTTCTGTGAGACCTTTTCGAATATCTTCATGAACAGTAG---TGTTCG 8607
8551 ----- 8607

Db 599 -----LysPheGluAspSerIlysalATyrLeuArgSerIlyserIlyGlnCysSer 615
QY 8608 AATTATTAATTTATGATAAATGATGGTTTCCAAATATACGAACATATGCTTCGAGAA 8667
616 AsnIleGluPheAsnsp-----GluThrPheTrpPhe----- 626
QY 8668 ACACCAAAAAGTTATAAAGAGCTTGCACT---TGTCACCTACTCTCTAAGAT----- 8718
627 ---ProAsnIlystrIyrgluIalacysMetValCysGluAsnProSerSerIlysalA 645
QY 8719 -----CCATGGATATAT-----TGTCACGAGTCAAAACAAGATGATTAAGCA 8766
646 LeuIlyProlIleYstrhrAsnValPheProlIleGluIlyserIlyIys-----SerGln 663
QY 8767 TTACAAACTTTTACCTCTCTGCGTGAAGAAATGATTAATGATATCTGATTAATGGAAC 8826
664 LeuSerSerIleuThrAspIlyserIlyAsn----- 673
QY 8827 GCATACCTGTTCTTAAATATCTTACATGATATMACAAAGTGTATGATTCCTCCAGAGA 8886
674 -----ThrProAsnSerSerGlyGly----- 681
QY 8887 AGACATTTATGTACAAACCTATCCTGCATATATATTAAGAAAGTGTATAAGAAAT 8946
682 -----AsnTy-----GlyAspArgGlnIle 688
QY 8947 TTAATAAA-----AACTCTTACTCTCTGCTTCAGTCA 8982
689 SerIlyArgspAspValHisIlyAspGlyProlYsgluValIlysserGlyIlyGln 708
QY 8983 GGACAAATTTGAGTCAAAATTAATAATCGAAGAGAGTTGTC----- 9027
709 ValProlYsIleAspAlaIleValIlystrGluAsnIleuPheTrpSerAsnArgAsnAsp 728
QY 9028 TTTGAGGCATGAATAATATGATATGATATGATATGATATTAAGAACTGATATG 9087
729 IleGluIlyIyglYstrIyserIyglYAspHisSerProValHisSerIlyAspIle 748
QY 9088 ATGACACTTCATTATCTGAAAAATTAATAATATTTGAACATCAAAATGAAGCAAC 9147
749 -----LysAsnIleuIleProlInArgValIleValSer 758
QY 9148 GAAATTCGTAAACATGCTGGGAAATTAATAGACGTACAGATAGCACCTATGTTATG 9207
759 GluAsn----- 760
QY 9208 GATATATAATTCGTACTTCAAAAGTACATTAAGTGAAGATGCTCATTAACCAAG 9267
761 -----LeuProlYs 763
QY 9268 GATGAAGAAATCAATCACTTCTCTGCTGTTAATGATGGCAAGCAAGCATGTAAG 9327
764 IleGlu----- 765
QY 9328 GAAAGAAACATGTAAGTATCATTAATAAACAATATGCTCTGTTCAAGAGATAT 9387
766 GlulysMetGluSerSerAspSerIleProlIle-----ThrHis 778
QY 9388 TTTGAAGCGTCAGAAATTAAGACAAACCTGATGTCAGATGATATTAAGAAAAATATAT 9447
779 IleGluIleGlu----- 782
QY 9448 AGCTTGAATATATGATTAATAAATACAAATGAAATCTAATATTAATTAAGCAATTA 9507
782 ----- 782
QY 9508 AAGATCAATCTTCAGTAAATATAGACATTAACCATCT----- 9546
783 LysGlyIlyInserSerAsnSerSerAspAsnAspProlValIleSerGlyArgGluSer 802
QY 9547 -----GAAGAAATGTTCAGTCATATATTAATA 9573
803 LysAspValAsnIleuHisTrpSerGluArgIleIyserGluAsnGluIlyValIleIys 822
```

```

QY 9574 TCAGAGATCTCAATCGCCTTGAGTTA----- 9603
Db 823 ThrspgluSerSerlySerllegluSerlyleProseraspGlnasnHis 842
QY 9604 AATGATATAATGAAATGTACAGAGAAACAAAATAATGAAATATGAA----- 9654
Db 843 SeraspLeuSerGlnasnAlaasnGlnaspSerasnGlnLysGluThrIleasn 862
QY 9655 -----TTCAGAGAGATCTCAAAATAATATATATCTGGTTATATTTGTTGAGAT 9705
Db 863 ProProserThrGluLysAsnLeuLysGluIleHis----- 874
QY 9706 GAAACACAAATAATCTACTGATGAAATATAAAGAGAGACAAAGTTGCT 9765
Db 875 -----TyrlysthrSeraspSeraspHisGlySerlyleLysSerGluIleGlu 892
QY 9766 CCTAAGACTCTATTTCTTACACCCCATGTAGATCTTTCATCAAGACCTTATTC 9825
Db 893 ProLysGluLeuThrGluGluSerProLeuThrAspLys----- 905
QY 9826 TCACACATGAGTAGACACATATGATCCTAAATAATGATATTTGAAAGTACT----- 9879
Db 906 LysThrGluSerAlaIleGlyAspLysAsnHisGluSerValLysSerAlaAspIle 925
QY 9880 -----ATCTCTGTGTTATGTA 9897
Db 926 PheGlnSerGluIleHisasnSeraspAsnArgAspArgIleValSerGluSerVala 945
QY 9898 -----TCGGCGTTAGGTTTGTGATAGCGCTTCATTTGAGAAATAATTC 9942
Db 946 GlnaspSerSerGlySerSerMetSerThrGluSerIleArgThrAspAsnLysAspIle 965
QY 9943 AATGCTGTGTGAGCTGTTGGTATACGATATCCCAAGAGAGAGATGGAATGCT 10002
Db 966 LysThrSerGluAspIleAlaProSerIleasn----- 976
QY 10003 AGCTTGGAATCCAAAAATAGTACATACATATAGAGTGTCCATATTAAGCAAAACA 10062
Db 977 -----GlyHisGlu 979
QY 10063 TATATATATATGAGAGAGATCTAGTGAATGAATGAATATATGAGGACTTATCT 10122
Db 980 LysIleGlySerSerAlaAspAspArgIleGlySerGluAspLysSerIleIleAspLysAsp 999
QY 10123 TCCCTGTGATATCTATCCGAAGATGATGAAGATGATATATATATATAT 10182
Db 1000 SerGluAsnPheGluAsnAsnLysSerSerHisSerAspIleLysGlnSerAsp----- 1017
10183 GTACCAAGTAGCTCTAAATATATAAATGATGATAGAGTAGTACAGACATCAAAAAGG 10242
Db 1018 AsnGluGlySerThrAspLysGluSerLeuThrGlu-----GluSerProLysGly 1034
QY 10243 GATATACCAAGT-----GATGATACCAAGATATGATATCA 10278
Db 1035 AspleuGluSerValSerProSerSerIleAspMetAspLeuLysProAsnLysSerSer 1054
QY 10279 CCACGTACCAAT 10290
Db 1055 ProValThrSer 1058

```

RESULT 4

US-10-153-273-8

Sequence 8, Application US/10153273

Patent No. US20020169305A1

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

Chilniss, Chetan

Miller, Louis H.

Peterson, David S.

Su, Xin-zhaun

Welliams, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

```

AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSER: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153, 273
FILING DATE: 21-May-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: Seq ID NO: 8:
US-10-153-273-8
Alignment Scores:
Pred. No.: 1,35e-33 Length: 921
Score: 649.50 Matches: 250
Percent Similarity: 39.48% Conservative: 127
Best Local Similarity: 26.18% Mismatches: 280
Query Match: 3.34% Indels: 288
Gaps: 48
US-10-087-013-1 (1-10628) x US-10-153-273-8 (1-921)
QY 2842 GATATATATCGAGAGAGATCTCTGGAAGAAAGAGGATGTAAGCTGCAAGCA 2901
Db 80 AspTylIleThrAlaSeraspIle-----GluAsnGlyGlyAsnSerIle----- 94
QY 2902 CATTTGGAAGACTTTTGTGTAATATACATTAAGTCACTCAAGCAAGCAAGAAATGATAAA 2961
Db 95 -----GlyAsnIleAspMetValIleSerAspLysAspAlaAsnGly 108
QY 2962 TATAT-----GATGATGCCCAATATTTAAATTTAGGCAAAATTTGG 3006
Db 109 PheAsnGlyLeuAspAlaCysGlySerAlaAsnIlePheLysGlyIleArgLysGluGln 128
QY 3007 TGGGAAGCTAATAGAGCCAAAGTAGGAGAACCATGAATGATATTAATTTGAAG 3066
Db 129 TrpLysCys-----AlaLysVal-----CysGlyLeuAspValCysGly 141
QY 3067 GATTAATGGGACACACATCAACACAAAGTACTTATTCGGATATAGTATACATACACA 3126
Db 142 LeuLysAsnGlyAsn-----GlySerIleAspLysAspGln 153
QY 3127 TTGATGATTAATCCCAAAAATTAAGATGATGATGACCAATGGGCAATGATGATGTC 3186

```


OY 4990 ----- 4990
 Db 824 rIleValleuAlaLeuGlySerIleAlaPheLeuPheMetLysValIleTyrIleTyrVa 844
 OY 4990 ----- 4990
 Db 844 lTpyrIleTyrMetLeuGlyValGlyAlaLeuSphTrIleTyrIleCysGlyCysIleCys 864
 OY 4991 -ANGTCCTTTTCTCTCTGTCGAC-----TACGTATATGTTTTCATGCA 5034
 Db 864 sIleCysIlePheIleCysValSerValTyrValCysValTyrValTyrValPheLeuTyr 884
 OY 5035 T-----TGATGCAATTTACACATCCAGAGATTAAAGATGAAT 5076
 Db 884 rMetCysValPheTyrIleTyrPheIleTyrValPheIleLeuLysMetLysLys 904
 OY 5077 GGCTTGCAGAAAGATGATGAGATGGCGACAGAGAGGTACAAATTGGCATAC 5136
 Db 904 s-----MetLysLys 907
 5137 TACAAAGAAAAAGAAAAAGAAAAATTAATAAAGCTGCGATG 5179
 907 smetLysLysMetLysLysMetLysLysAlGlyLysAlGly 921
 RESULT 5
 US-10-153-273-4
 Sequence 4, Application US/10153273
 Patent No. US20020169305A1
 GENERAL INFORMATION:
 APPLICANT: Sim. Kim I.
 Chitnis, Chetan
 Miller, Louis H.
 Peterson, David S.
 Su, Xin-zhaun
 Wellens, Thomas E.
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knodde Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/153,273
 FILING DATE: 21-May-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/210,288
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Fuller, Michael
 REGISTRATION NUMBER: 36,516
 REFERENCE/DOCKET NUMBER: NIH121.1fmdv1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1435 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEetical: NO
 ORIGINAL SOURCE:

ORGANISM: Plasmodium falciparum
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:
 US-10-153-273-4
 Alignment Scores:
 Pred. No.: 1,566-31 Length: 1435
 Score: 618.00 Matches: 339
 Percent Similarity: 36.808 Conservative: 223
 Best Local Similarity: 22.208 Mismatches: 538
 Query Match: 3.188 Indels: 428
 Gaps: 72
 US-10-087-013-1 (1-10628) x US-10-153-273-4 (1-1435)
 OY 3301 GAAGCTTGTATGATTAATTAATGATTAATAGATTTANGAAGAAACAA-----TGG 3351
 Db 20 LysAlaArgAsnGluTyr-----AspIle-----LysGluAsnGluLysPheLeu 34
 OY 3352 AATTAATTAATTCAGATTAATTAACAAAGATTAATGACAGACAAATGCTGTAGTAAT 3411
 Db 35 AspValTyrLysGluLysPheAsnGluLeuAsnLysLysTyrGlyAsnValGluLys 54
 OY 3412 AGTGTATTGAAGCTTCAGTACTGCCAAAAATCATATATGACAGAGATGTTATGATTT 3471
 Db 55 Thr-----AspLysLysIlePheThrPhe 62
 OY 3472 TTG---TCGGAATTAATACAAATAATGTGCGAAAGTAATTAAGT----- 3516
 Db 63 IleGluAsnLysLeuAsnIleLeuAsnAsnSerLysPheAsnLysArgTyrLysSerTyr 82
 OY 3517 GGTACTAGTGTGAAGTAACTGCTGATTCGTACTGACACACAGATGAAATGTTGAGACA 3576
 Db 83 GlyThrProAspAsn-----IleAspLysAsnMetSerLeuLeu----- 96
 OY 3577 TATTCATATACAGAAATTTGATGATTTGATGATTCATGACATCAAAAGAGTTTGATGAA 3636
 Db 97 ---LysHisAsn-AsnGlu---GluMetPheAsnAsnAsnTyrGlnSerPheLeuSerTh 114
 OY 3637 AAAAGATGATGATGATTAACGAAATAATATGCTTTAGATTAACACAGAGACATGAT 3696
 Db 114 rSerSerLeuIleLysGluAsnLysTyrValProIle-----AsnAlaValAlaGlyVa 131
 OY 3697 GGTGCTGTGCTGTGTAAGAGTGAATGACGAAACCGACAGGATACAGATTAACCAAAAA 3756
 Db 131 lSerArgIleLeuSerPheLeuAsp----- 139
 OY 3757 AAAGCGAAGAAAGATACGGAATGTAACAGTG-----AATGATATACTTAA 3806
 Db 140 -----SerAlGlyLeuAsnAsnGlyArgAsnThrSerSerAsnAsnGluValLeuSe 156
 OY 3807 AGAA---AACGATGGAAGAAACAAAGTAGAA---GATTGATCAGCAAAAGATAGTA 3860
 Db 156 rAsnGlyArgGluLysArgLysGlyMetLysTyrPheCysLysLysLysAsnAspArgSe 176
 OY 3861 TGCATATCCCGATTGGCAATCGCAAAATTAATTAAGTGAAGACCTTCGTGTAT 3920
 Db 176 rAsnTyr-----ValCysIle 181
 OY 3921 GCCCCTTAAGAAAGAAAGTATGCTGATCTTGGCAATGATTAAGATTAAGATTAAGAA 3980
 Db 181 eProAspArgArgIleGlnLeuGlyIleValAsnLeuSer-----IleIleIle 197
 OY 3981 ATTCAATCAACAGTATTAATTAAGAAAGCTTTCATCAATGCGACGACAGAAACATT 4040
 Db 197 sThrTyrThrLysGluThrMetLysAspHisPheIleGluAlaSerLysGlu----- 215
 OY 4041 CTTCATGCTATTAATTAATTAAGAGTGTGAAGAGTGAAGAAATGAACCTGCATTAAGAA 4100
 Db 216 ---SerGlnLeuLeuLeuLysLysAsnAspAsnLysTyrAsnSer----- 229
 OY 4101 AAAAGAAAGCAAAATTCCTCCGCAATTTTGGATTCATGATTCATGATTTGGAGATTA 4160
 Db 230 -----LysPheCysAsnAspLeuLysAsnSerPheLeuAspTyr 242

[illegible][illegible]

```

Db      810  urhraspaalalyasprhserasnleuylsleuylslyaspvalaspalleseme 830
OY      6246  GGTAAAGCCATCTTTCGCAATTAACCAATAGCAGGGGTGAATAGGGTGTATATCC 6305
Db      830  tProlysalvalalleglyserseproasn-----Aspsnllseasn 844
OY      6306  AAAAAGCTATGCACATATCCATAATGGGGTGTATAGTAACTTAAGAAATGA 6365
Db      845  -----ValThrgluIngluYaspsnll 852
OY      6366  AATGGCATATGTATGCTCTAGAGAAAAAATATATGTATTAATATATACATATTT 6425
Db      852  eserglyval-----AsnserlyspProleuseraspsvalatrgproas 867
OY      6426  AATATTAAGAACTGAATAATAGCGTGCACATGTATATAAGAGCGCTTTTAAATGTC 6485
Db      867  pulysgluleugluaspclnasnseraspsglusergluIngluThrValVal----- 884
OY      6486  AGCAATAGAACTCAATTTTGTGTTAAATATATATATATATATATATATATATAT 6545
Db      885  -----AsnhisIleSerlySerProserIleasn 895
OY      6546  TGAATTCGAAAAATGGAACAATTCAGATGAATTTAAAGAAATATATATATATATG 6605
Db      895  ncllyaspspserglyser-----GlySeratIatThrValse 907
OY      6606  TGATTTATTAAGATATGTTTTTGGACATGATTTCTAATGATATAAAAAATTAATCT 6665
Db      907  rgluserseaserasnthrglyleuserIleaspsapargasnnglyAspThrPh 927
OY      6666  AACAAATAGTGAACAAACATTCATGAATAATATATAGAAATTAAGAAATTAAGAA 6725
Db      927  eValarghrhglnasprhthlaasnThgluasPallleatglysgluasAlaAsp 947
OY      6726  AGATGAAGAAATACGTAAATATTTGGAGAGAAAAATTAATTAATTTGGAGAGAT 6785
Db      947  saspsgluaspsgluYsglyAlaaspgluIngr----- 958
OY      6786  GATATATGATTAATCTTATCATCTACAGACAGCAAAC-----GAAAAAGAAA 6833
Db      959  -----HisSerThrserGluserseaserProgluIngluYse 972
OY      6834  AATAGAGATATATACAGTACATGATGACCAAACTGACGCGCTTCCCTGAAGATT 6893
Db      972  tleuthraspsnngluYglyAsnserleuasn-----Hisglu 986
OY      6894  TGTAAAGAGCCCAATTTTGAATGGTTCACAGAAATGGCAGAGAAATTTGTATTA 6953
Db      986  uVallysgluThrsrserasnseraspsnValgluInglusergluYllevalasme 1006
OY      6954  GAGGAAGAGACAGTTGCTTAAAA-----TTGAGAGCGGG 6986
Db      1006  tAsnValgluYsgluYsaspThrLeugluasnProserseaserleuasnpglu 1026
OY      6987  CTGTAAGCAATATAGTGTATGTAAT----- 7017
Db      1026  YlysalahisgluInglusergluProasnleuserseaspsgluaspsmetseasn 1046
OY      7018  -----GACGTAAGACACAAAGATGTCAGAGCGCTGTACATATCAAA 7064
Db      1046  rProglYProleuaspsnThrsrsergluIngluThrThrgluYr-----IleSerAsn 1064
OY      7065  TTTTATTAAGAACTGAATAATGATATGAAGACAAAGAAAGTTCAAAAGAT-- 7122
Db      1064  ngluThlysalasnngluatrgluaspgluThrleuthrlysgluYrGuasPll 1084
OY      7123  -----AAAGATGGCAAAAGTATAGAGATTATCC 7151
Db      1084  eValleuYserHismetasnargluseraspsapgluIngluYrAspGu----- 1102
OY      7152  TTTTACGAAAGACATAGAGAGCAACATGCTGCTCATGATATTTAAACATGAAT 7211
Db      7211  ----- 7211

```

```

Db      1103  -----AsnseraspleuserThrValasnaspsglusergluaspaalagluAlasme 1120
OY      7212  AAAAGATATATGCGCAATTAAGATTTGTTGTATGCAAAAAACCTTCTCAACATACC 7271
Db      1120  tlys-----Glyasn-----AspThrserGlumetserHisasnserseasin 1136
OY      7272  AAAAACAACAACAATATACATATCCGATTCGATATATATATATATATATATATAT 7329
Db      1136  egluseraspsgluIngluY-----AsnspmetlysthrValGlyaspLe 1151
OY      7330  -----TATGTCCTGAGAAATTTAACAAGTGTGCTCTCACTTCAAAA-- 7377
Db      1151  uglYThrThrsValgluasngluIleSerValProvalThrgluIngluIleaspglu 1171
OY      7378  -----AAGGATCTATGATTCATACAAAAAATTAATCTGAACCTTAATATACC 7424
Db      1171  sleuYrgluserlysgluserlyIleHis-----lysalaIngluIngluYrleuse 1189
OY      7425  TATGAATGTGTAGAGACAGACATATTTATCTTAAGACAGCAAAATATATATGGA 7484
Db      1189  rHisThraspleHislys-----IleasnProgluaspsargasnserasnTh 1205
OY      7485  TATTACTTGAAGAAAAATTTATATACCTATTAAGTGTACAAAGGAAAAAGAACTAAA 7544
Db      1205  rleuHisleuYsasps-----IleargasnngluIngluasngluar 1218
OY      7545  TAGTTGACATAATATAT 7563
Db      1218  ghIsleuthrasnngluInsn 1224

```

```

RESULT 6
US-09-924-154-17
; Sequence 17, Application US/09924154
; Patent No. US20020127241A1
; GENERAL INFORMATION:
; APPLICANT: Natum, David L.
; APPLICANT: Slim, Kim L.
; TITLE OF INVENTION: Anti-plasmodium Compositions and Methods of Use
; FILE REFERENCE: 05213-0465 43170-262105
; CURRENT APPLICATION NUMBER: US/09/924,154
; PRIOR FILING DATE: 2001-08-07
; PRIOR APPLICATION NUMBER: US 60/223,525
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 1501
; TYPE: PR
; ORGANISM: Mammalian
US-09-924-154-17

```

```

Alignment Scores:
Pred. No.:      8,01e-31      Length:      1501
Score:          607.00      Matches:      345
Percent Similarity: 34.378      Conservative: 217
Best Local Similarity: 21.108      Mismatches: 548
Query Match:      3.12%      Indels:      525
DB:              10      Gaps:      71

```

```

US-10-087-013-1 (1-10628) x US-09-924-154-17 (1-1501)
OY      6153  AATGCGAGCAACAACGAAACCCGACGACGACACAAACAGAAACGAAACATCAAC 6212
Db      42  AsnValgluThrasnAsn-----AsnAsnAsnAsnAsnAsnAsnAsn 58
OY      6213  AGCAACAACAACAATCTGACGTGGGACAAATGTAAGGCGCATCTTTCGAAATTAACC 6272
Db      59  SerasnaspAlameSerPheValasngluValIleatrgPheIlegluasngluYsas 78
OY      6273  AGAT----- 6276
Db      78  paspLysgluaspLylyVallyIleIleSerarGProValgluasnthrleuHisar 98

```


QY 6276 ----- 6276
 Db 98 gtyProValSerSerPheLeuAsnIleLysIleTyrglyArgLysGlyIleuTyreAs 118
 QY 6277 ----- AGCAGGGGTGGACTAGAGGGTGTGTAATCCAAAACGATGACGA 6320
 Db 118 nArgAsnSerPheValGlnArgSerTyrlleArgGlyCysGlyLysArg-----Se 136
 QY 6321 AATATCCAAATGGGGTGTGATGTAGGTAACTAAAGAAAATGAAAGCATGTATGAT 6380
 Db 136 rthrHstHtrPleCys-----GlnsInLysGlyAsnAsnAsnIleCysIle 152
 QY 6381 GCCCTCTAGAGAAAAAATATGTAATATATATATATATATATATATATATATATATAT 6437
 Db 152 eProAspArgValGlnLeuCysIlePheValLeuGlnAsnLeuLysAsnSerGlyse 172
 QY 6438 TCAAAATAGCGTGAACATGATATATATATATATATATATATATATATATATATATAT 6497
 Db 172 rgluHtrHtrAspArgLysLeuLeuArgAspLysValPheAspSerAlaMetTyrglyIu 192
 QY 6498 TCAATTTTGTGGTAAAT 6557
 Db 192 rAspLeuLeuTrpAsnLysTyrglyPheArgIle----- 203
 QY 6558 TGGACAAATCCAGATGATTTAAAGATATATATATATATATATATATATATATATATAT 6617
 Db 204 -----PheAspAspPheCysAspAspValLysAsnSerTyrlleuAspIlyLysAs 220
 QY 6618 TATGTTTTTGGACATGATTTCTATATATATATATATATATATATATATATATATATATAT 6677
 Db 220 pValIlePheGlyTrpAsnLeu-----AspLys-----AsnAsnIle 232
 QY 6678 AACAAACATTCATGATGAAAT 6737
 Db 232 eSerLysLeuValGlnGluSer-----LeuLysArgPhePheLysLysAspSerSerVa 250
 QY 6738 ACGTAAA-----ATATTTGGGAGAAAAATATATATATATATATATATATATATATAT 6791
 Db 250 IleuAsnProThrAlaTrpTrpArgTyrglyTrpArgLysTrpMetIleG 270
 QY 6792 TGGATTAACTTATCATCTC-----ACAGAGCAAAACGAAAAAGAAAAATTTAG 6839
 Db 270 nProTyrlaIle-----HisLeuGlyCysArgLysProAspGluAsnGlu----- 284
 QY 6840 AGATATATTCAGTACATGACATGACATGACATGACATGACATGACATGACATGACATG 6899
 Db 284 ----- 284
 QY 6900 AAGCCCCCAATTTTGAATGTTTCAACAGATGCGCAGAGAAATTTTGTAT-----AA 6953
 Db 285 ---ProGlnIleAsnArgTrpIleLeuGlnIuTrpGlyLysTyraSncysArgLeuMetly 303
 QY 6954 GAGGAAGCAAGTGTGTAAATTTGAGGCGGGCTGT-----AAGCAATATGAGTGA 7007
 Db 303 sGluLysGluLysLeuLeuThrGlyGlyCysSerValAsnArgLysLysSerAspCysSe 323
 QY 7008 TGGTATATGACCGGTAGACACAAACATGTCAGAGCGGTGTATCATCATCATCATCATCAT 7067
 Db 323 rthrGly-----CysAsnAsnGlnCysTyrlleTyrlleArgSerle 336
 QY 7068 TATTAAGAAAGTGGAAATGCAATATGAAAGCAAGAAAGAAAGTGAAGATTAAGA 7127
 Db 336 uIle-----AsnArgGlnArgTyrglyValSerIlele 347
 QY 7128 TGGCAAAAGATAT-----AAGGATTATCCTTACTAGAAAGACATAGAGAGG 7178
 Db 347 uGlyLysLysTyrlleLysValAlaTrpTyrllePheArgArgLysIleValGlnPr 367
 QY 7179 AACATGTCTCATGAT 7238
 Db 367 oAspAsnAlaLeuAspPheLeuLysLeuAsnGlySerGly----- 380
 QY 7239 TTTCTGTATGCAAAACCTTTCTTACACACTACCAAAACACAAACATCATCATCATCATCAT 7298

Db 381 ----Cys----- 381
 QY 7299 CGATGTAAATGATATGCAAGAAATCGGTGATTTGTTCTGCAAGATTTAAAGATGTGA 7358
 Db 382 -----LysAspIleAspPheLysPro-----PhePheGlnPheG 393
 QY 7359 GTCTCCGTAATTTCAAAAAGGATCTATGATTCATACAAAAAATTAATTAATTAATTAATTA 7418
 Db 393 uTyrglyLysTyrglyGlnLys----- 400
 QY 7419 AATACCTATGATTTGTGTAGACAAAGCAGACATTTATTTATCTTAAAGAACAGAAATTA 7478
 Db 401 -----CysMetCysGlnSerTyrlleAspLeuLysIleGlnPheLysAsnAs 416
 QY 7479 TATGATATATTCCTTGAAGAAAAATTTATATCTATATGATGCTACAAAGAAAGAAAG 7538
 Db 416 nAspIleCysSerPheAsnAlaGlnThrAspThrValSerSerAspLysArgPhe----- 434
 QY 7539 TAAAAATAGTTGGACTAT 7598
 Db 435 -----CysLeuGlnLysLysGlnPheLysPro----- 443
 QY 7599 ATATATAGACAGAAACCCCTTGTGAAATATAGAGAAAGAAATCGTTTAAAGTATATTA 7658
 Db 443 ----- 443
 QY 7659 TGAATGAATGTTACAAAAATTCAAAGTTCTATCAGAGAAAAAAGATATGTATAC 7718
 Db 444 -----TrpLysCysAspLysAsnSerPheGlnuTrpValHisHisLysGlyValLysValSe 462
 QY 7719 TCCAGAGAGAACATATATGCTTAAAGAAATTTAGATGAAT-----AAATTTGA 7769
 Db 462 rProArgArgGlnGlyPheCysLeuGlnLysLeuAsnTyrlleLeuAsnAspIlyle 482
 QY 7770 AAGACTTAAGAT 7829
 Db 482 rAsnValHisAsnSerGlnLeuLeuIleGlnIlele-----MetAlaSerLysGlnGln 501
 QY 7830 ATATAGCAATATATAAAAAC-----TTCACCTGACAGACGGGCGCCATATGA 7877
 Db 501 yLysLeuLeuTrpLysLysHisGlyThrIleLeuAspAsnGlnAsnAlaCys----- 518
 QY 7878 TCCAAATATGTATATATATATATAT-----AGTTTCGCTGATCTGGGTGACATATG 7928
 Db 519 -----LysTyrlleAsnAspSerTyrlleValAspTyrlleLysAspIleVa 532
 QY 7929 TAGAGAACATATATGTTACGAATTTGTTACTTACCTCCCGTACGAATTAATTAATTA 7988
 Db 532 lIleGlyAsnAspLeuTrpAsnAspAsnSerIle-----LysValGlnAsnAsnLeuAs 551
 QY 7989 TAAAGTTTGAATACATATATATGGAAGAAATGGAAGAAATTAAGATGAATTAAG 8046
 Db 551 nLeuIlePheGlnArgAsnPheLys-----TyrlleValGlyAlaTrpAsnLysLe 567
 QY 8047 -TACAAAGATGTACAAACGTTTCTGCTGCTGGTGGAGATGCTATATAGAAAGATTTTGG 8105
 Db 567 uPheLysTrpIleLysGlnLeuLysAsnValTrpTrpIleLeuAsnArgLysLysValTr 587
 QY 8106 GAAAGCAATGACGTGCAAAAGCACCAGAAAGTGCAAAACCTTTTAAAGAAAGAAATGGA 8165
 Db 587 rGlnSerMetArgCys----- 592
 QY 8166 TGGATTGAAGCGCAT 8225
 Db 593 -GlyIleAspGlnValAspGlnArgArgLysTrpCysGlnuArgIleAspIleu----- 610
 QY 8226 TGATGAT 8285
 Db 611 -----GluAsnMetProGlnPhePheArgTrpPheSerGlnTrpAlaHisPhePheCysLy 629
 QY 8286 ASCAGTGAATGGAAGAAATTTGAAATAATTTAAATAATGATGATGATGATGATGATGATGAT 8345

Db	915	-----ThrGluGluThrProSerIleSerGluGluProLysGlyAsnGluGluLysLysArg	933
Oy	6726	AGATGAAGAA-----TTACGTAATATTTTGGGAGAAAAATTAATAATTATTGGGAAAG	6782
Db	933	gaspaspaspserleuserlysile-----	941
Oy	6783	AATGATATATGAGTTAACTATTCATCTCCACAGCCAAACGAAAAAGAAAAATTTAGCA	6842
Db	942	-----SerAlaIleValAla-----	953
Oy	6843	TAATTAACGATACAAATGACATGACCAACTGACAGCCCTTCCTCTGGAAGGATTGTAAAG	6902
Db	953	pAllyAspThrSerAsnIleuLysLeuGlyAspValAsp-----IleSerIe	971
Oy	6903	GGCCCAATTTTGGATGGTTCACAGAAATGGCGAGAAATTTTGTAAATAGAGAGCA	6962
Db	971	tProLysAlaVal-----IleGlySerProAsnAsp	982
Oy	6963	ACAGTTGTTAAATTTGGAGCGGGCTGAAAGCAATATGAGTAAATGCTAGTAAAGCGG	7022
Db	982	pAsnIleAsnValAlaThrGluGlnGly-----AspAsnIleSerGlyValAsnSer--	998
Oy	7023	TAAAGACAAAGAAATGTGCAGAGCGGTGTAAACATATCAAAATTTTATTAGAACTGAA	7082
Db	998	-----	998
Oy	7083	AACTGAATATAGAAAGACAAAGAAAGTTCAAAAAGTAAAGATGGCAAAAGTATTA	7142
Db	999	-----LysProIleuSerAspAspValAlaArgProAspLysAsnIleGluValLys	1015
Oy	7143	GGATTATCTTTCTACTGCAAGAGACATAGAGAAAGCAACATGTGCTCATGAATATTAA	7202
Db	1015	sgLInHISThrSerAsnSerAspAsnValGlnGln-----SerGlyGlyIleValAs	1032
Oy	7203	CATGAATTA-----AAAGAAATTATGTGGCAATAGATGTCTTCTATGCAAAACCTTC	7259
Db	1032	mMetAsnValAlaGluSerIleu-----LysAsp-----ThrLeuGluAsnProSe	1047
Oy	7260	TTCAACATACCAAAAACAAACACACACACATCACATATCCGATGCTAAAT-----	7308
Db	1047	rSerSerLeuAspGluLysAlaIleAsnGluGluIleuSerGluProAsnIleuSerSerAs	1067
Oy	7309	-----GATATG-----CCAGATCGCTGATTAATGTCTCTGAGAAATTTAACAACTG	7355
Db	1067	pGlnAspMetSerAsnThrProGluProLeuAspAsnThrSerGlu-----	1083
Oy	7356	TGATGTCCCTGAAGCTTCCAAAAAGGAGTCTATGATTCATACAAAAAAATTAATCAAC	7415
Db	1084	-----ThrThrGluAlaGlyIleSerAsnAs	1091
Oy	7416	TAAATATACCTATGATGTGTAGAGAAAGCAGCATATTTATTATCAAAAGAGCAAAA	7475
Db	1091	nGluTyrLysValAsn-----GluAlaGlyGlu	1100
Oy	7476	TAATATGGAATTAATCTTGAAGAAAAATTAATTAATCTATTAAGCTTACCAAGAAAAAGA	7535
Db	1100	yGluArgThrLeuThrLysGluTyrGluAspIleValIleuLysSerHisMetAsnArgG	1120
Oy	7536	AAGTAAAT-----AGTGGACATTAATTAATCC	7565
Db	1120	uSerAspAspLysLysIleuTyrAspGluAsnSerAspLeuSerThrValAsnAspGluSe	1140
Oy	7566	TTTGATTCCTTAAG-----AAACCTATGACCTGTAAATATATAGAGAAAGAAACCTTG	7622
Db	1140	rGluAspAlaGluAlaLysMetLysGlyAsnAspThrSerGluMetSerHisAsnSerSe	1160
Oy	7623	TGAATAATGACAAAGAAATGTTTAAGCTAGATTATGAATGGAATGTTCACAAAATTC	7682
Db	1160	rGlnHisIleGluSerAspGlnGluIleAsnAsp-----MetLysThrValGlyAspLe	1178
Oy	7683	AAAGTTCATCAGCAAAAAAGAGATGTCTGACTCCAAAGACAGAA--CATATGTG	7739

```

Db 1178 uG1yThrHisValGlnsngluIleSerValProValThnGluIuIleasrGlu 1198
Oy 7740 CTTAGGAAATTAGATAAATTAAAT-----GAAAGACT----- 7776
Db 1198 sleuArGluSerLysGluSerLysIleHisLysAlaGluIuIuArgLysSerHisTh 1218
Oy 7777 -----AAGATTAATAATTATTC-----CTTAATAATGCT 7805
Db 1218 rasPLeHisLysIleasnProGluAspArgAsnSerAsnThLeuHisLysAspDl 1238
Oy 7806 TCGT-----CGAATCGACGAAATGAGAAAGACATATATATAAAACT 7850
Db 1238 earGAsnGluIuAsnGluArgHisLsleuThArgAsnGlnHisIle-----AsnI 1254
Oy 7851 CAATCAGAGAACGGGTGCCAATGAATCCAAATATGTGTACTAGTAATAATAGTTTGC 7910
Db 1254 eSerGluGluArg-----AsrLsGluIuLysHisGluPheHis 1266
Oy 7911 TGATCTG-----GGTACATAGT----- 7929
Db 1266 sThMetLAsnAsnLeuHisGluLysArgLysValSerGluArgSerGlnIleasnHisSerH 1286
Oy 7930 -AGAGACAGATATGTTGAATGATGGTGTACTTAACCTCCCGTAGAAATAAATTATA 7988
Db 1286 shISGluYAsnArgGlnAspArgGluGluYAsn----- 1296
Oy 7989 TAAGTTTTTGAAATACATATATATGAAAAATGAGAAATTAATAAGGTAGAAATATA 8048
Db 1297 -----serGluYAsnValLysAsnMetLArgSerAsnAsnAsnAsnAsnPh 1310
Oy 8049 CAACGAGTGTACAAACGTTTGGTTCTGCTTGCTGGTGGATGCTAATACAAAAGATATTGGAA 8108
Db 1310 eaShAsnIle----- 1313
Oy 8109 AGCAATGACGCGAAGACACACAGAGATGCAAACTTTT-----AGAAAAGAAATGGA 8165
Db 1314 -----ProSerArgLysAsnLeuArgLysArgLysLysLeuAspLeuAs 1327
Oy 8166 TGGATTGAAAGCGCTAATTAATTAACAAGATAAGTGTGGACATTAAGAGCATCCACTGT 8225
Db 1327 PleuArgLys-----AsnArgAsnAspSerThTh 1337
Oy 8226 TGAATGATTATATACCTCAACGCGTTTGCATGATGACTGAATGGTGTGAATTAATTGAA 8285
Db 1337 rLysGluIuLeuIle-----Ly 1342
Oy 8286 AGCACTGATGGA-----GAATGCGAAAATTTAAAAATCATGTGATCAC- 8331
Db 1342 sLysLeuAlaGluIleAsnLysCysGluAsnGluIuIleSerValLysTrpCysAspHisMe 1362
Oy 8332 -----TGTAAACATCTGACAGATGCAAGATGATTAATGAGAA 8372
Db 1362 tLeHisGluIuIleProLeuYsTrp-----CysThnLysGluLysThArgAs 1379
Oy 8373 TAACTGCAACAGTGTAAAAAGAGATGTCAGAATATATAAAATTTGTTCTTAATGGA 8432
Db 1379 nLencuYsAlaValSerAspArgLysCysMetSerTrp----- 1391
Oy 8433 ATCTCTATTCGATATACAAATCAATAATC-----AAAGATTGTATGACA 8488
Db 1392 -----PheThrTrpAspSerGluIuArgLysTrpAsnCysThnLysArgGluPheAspAs 1409
Oy 8481 ACCAATATATACA 8493
Db 1409 pProSerTrpThr 1413

RESULT 8
US-09-924-154-14
: Sequence 14, Application US/09924154
: Patent No. US20020127241A1
: GENERAL INFORMATION:
: APPLICANT: Narum, David L.
: APPLICANT: Sim, Kim L.

```

RESULT 8
US-09-924-154-14
; Sequence 14, Application US/09924154
; Patent No. US2002012741A1
; GENERAL INFORMATION:
; APPLICANT: Narum, David L.
; APPLICANT: Sim, Kim L.

TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
 FILE REFERENCE: 05213-0465 43170-262105
 CURRENT APPLICATION NUMBER: US/09/924,154
 CURRENT FILING DATE: 2001-08-07
 PRIOR APPLICATION NUMBER: US 60/223,525
 PRIOR FILING DATE: 2000-08-07
 NUMBER OF SEQ ID NOS: 17
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 14
 LENGTH: 1143
 TYPE: PRF
 ORGANISM: Mammalian
 US-09-924-154-14

Alignment Scores:
 Pred. No.: 3.3e-30 Length: 1143
 Score: 597.00 Matches: 286
 Percent Similarity: 36.11% Conservative: 195
 Best Local Similarity: 21.47% Mismatches: 432
 Query Match: 3.07% Indels: 419
 Gaps: 59

US-10-087-013-1 (1-10628) x US-09-924-154-14 (1-1143)

QY 6006 TCAAGTGCACGAGGCCACACGCTAGCAAGGAAACCCGTCACCGGGTATTCAC 6065
 Db 45 SerAspIleSerArgValAsnSerProGluLeuAsnAsnHisLysThrAsnIleTyr 64
 QY 6066 GATATCAAAAGGACGCGCATCGAAAAAGAGCAAAACGCGCGCCCTACAAAAACACC 6125
 Db 65 AspSerAspTyrGluAspValAsnAsnLysLeuIleAsnSerPheValGluAsnLysSer 84
 QY 6126 GAAAAAGTGGAAATCT-----AACACAGAAATGCGAGCAAAACAGAAC 6173
 Db 85 ValLysLysArgSerLeuSerPheIleAsnAsnLysSerTyrLysAspIleIle 104
 QY 6174 CCGAGAGC-----AGC 6185
 Db 105 ProProSerTyrSerTyrArgAsnAspLysPheAsnSerLeuSerGluAsnGluAspAsn 124
 QY 6186 ACAACAAACAGCAAAACAGCAATCAACAGCAACAAACAGAAATCTGACGTGGCACAAT 6245
 Db 125 SerGluAsnThrAsnSerAsnAsnPheAlaAsnThrSerGluIleSerIleGlyLysAsp 144
 QY 6246 GGTAAAGCCATCTCTTGTGAATTAACAGATAGAGGGGTGATAGAGGTTTATCC 6305
 Db 145 AsnLysGlnTyrThrPhe-----IleGln-LysArgThrHisLeuPheAlaCysGlyI 162
 QY 6306 AAAAACGATGACAAATATCTTAAGAGGGGTGTATGATAGTCAAGTCAAGAAATGA 6365
 Db 162 eLysArg-----LysSerIleLysTrpIleCys-----ArgGluAsnSe 175
 QY 6366 AAATGGCATA-----TGATAGCTCTAGAGAGAAAAATATATGTTAATAATATATACA 6419
 Db 175 rGluLysIleIleThrValCysValProAspArgLysIleGlnLeuCysIleAlaAsn 195
 QY 6420 ATATTTAAATTAAGAACTGAATTAAGCGTCAATGATATATAAAGAGCTTTATTA 6479
 Db 195 uAsnSerArgLeuGlnIleThrMetGluLys-----PheLysGluIlePheLeuI 211
 QY 6480 ATGCGACCAATAGAACTCAATTTTGTGTTAAATATATATGAATCTGACGC 6539
 Db 211 eSerValAsnThrGlnAlaLysLeuLeuTyrAsnLysAsnGluGlyLysAspProSerI 231
 QY 6540 AGAA-----ATGAATTTGCAAAATGAGAACAAATTCAGATGATTTAAAGAAATATATATTA 6596
 Db 231 ePheCysAsnGluLeuArgAsn----- 238
 QY 6597 TACATATGCTGATTAATAGATATGTTTGTGAAGCTGATATTTCTATATGATTAATAAT 6656
 Db 239 -SerPheSerAspPheArgSerPheIleGlyAspAspMetAspPheGlyGlyAsn 258
 QY 6657 TATAACTGTATCAAAATAGTGTACAAACATCTTCATGAAATATATTAAGAAAAACAGGA 6716

Db 258 rAspArgValLysGlyTyrIle-----AsnLysLysPheSerAs 271
 QY 6717 TAAAAAAGATGAAAGATTTACGTAAATA-----TTTTGGAGAA 6758
 Db 271 PTrTTrTTrLysGluLysAsnValGluLysLeuAsnAsnIleLysLysGluTrpTrpGluL 291
 QY 6759 AAATAAAAATTTATTTGGAGAAAGATGATATGATTAATCACTATCATCTCACAGACA 6818
 Db 291 sAsnLysAlaAsnLeuTrpAsnHisMetIleVal----- 302
 QY 6819 AAAGAAAAAGAAAAATTAAGATTAATTTACCAATGACATGACAAACTGACGCC 6878
 Db 303 -AsnHisLysGluLysAsnIle-----SerLysGluCysAlaIleIleP 316
 QY 6879 TTCCCTTGAAAGATTTGTAAGAAAGCCCAATTTTGATGATGCTTCACAAATGGGACGA 6938
 Db 316 oAlaGluGlu-----ProGlnIleAsnLeuTrpIleLysGluTrpAsnG 331
 QY 6939 AGAATTTTGATTAAGAGGAAGAACAGTGTAAATTTGAGAGCGGCGTGT----- 6990
 Db 331 uAsnPheLeuMetGluLysLysArgLeuPheLeuAsnIleLysAspLysCysValGluAs 351
 QY 6991 -AAGAAATATGAG-----TGTAATGCTAGTATATAGACGTTAAACACAAAGATGTCAGAGCC 7046
 Db 351 nLysLysTyrGluAlaLysPheGlyGly-----CysArgLeuPr 364
 QY 7047 GTGTGTAATATCAAAATTTTATTAAGATGAAAGTAACTGATATGAAACAAAGAGA 7106
 Db 364 oCysSerSerTyrThrSerPheMetLysLysSerTyrThrGlnMetGluValLeuThrAs 384
 QY 7107 AAAGTCCAAAAAGATTAAGATGCAAAAGTAAAGATTAATCTTCTGACGAAAGAGA 7166
 Db 384 nLeuTyrLysLysLysAsnSerLys----- 392
 QY 7167 CATAGAGAACCAACATGCTCATGAATATTAACATGAATTAAGAAATTAATGTGC 7226
 Db 393 -ValAspLys-----AsnAsnPheLeuAsnAspLeuPheLysLys----- 405
 QY 7227 CAATTAAGATTTGTTCTTGTATGCAAAAACCTTCTTCAACTACCAAAAAACACACACA 7286
 Db 405 ----- 405
 QY 7287 ATCACAATATCCGATGATATGATATGCCAATGCTGATTAATGTTCTCGAAGAAAT 7346
 Db 406 -----AsnAsnLysAsnAsnAspLeuAspAspPhePheLysAsnGluLysGluTyrAs 422
 QY 7347 TAACAAGTGTGAGTCTCGAAGCTTCAAAAAAGGATCTATGATTCATACAAAAAAAT 7406
 Db 422 pAspLeuCysAspCysArg-----TyrThrAlaThrI 433
 QY 7407 TACTGAACCTAAATATACCTATGAATTTGTATAGAAAGCAGCATTTATTTATCTTAAGA 7466
 Db 433 eIleLysSer-----PheLeuAsnGlyPr 441
 QY 7467 AGCAGAAAAATTAATGATATTAATCTTGAAGAAAAATTTATACCTATTAAGTGTACAAA 7526
 Db 441 oAlaLysAsnAspValAspIleAla----- 449
 QY 7527 GAAAAAGAAAGTAAATATGTTGACCTAATTAATATATCTTGCATCTGACAAACCTTA 7586
 Db 450 -----SerGlnIleAsnValAsnAspLeu----- 457
 QY 7587 TGCACCTGATTAATTAATATAGAGAAAGAACCCCTGTGAATAATAGAGAAATCTGTT 7646
 Db 458 -----ArgGlyPheGlyCysAsnTyrLysSerAsnAsn----- 468
 QY 7647 TAACTAGATTAATGATGAGAAATGTTACAAAAATTCAAAGTTCATGACGAGAAAAAAG 7706
 Db 469 -----GluLysSerTyrAsnCys-----ThrGlyThrPheThrAsnLysPheProG 484
 QY 7707 AGTATGCTACCTCCAGAGAAAGAACATATGCTTAAAGAAATTA----- 7752

Db 484 ythCysGluProIaIrgGlnIthLeuCysLeuGlyArgThTyIleuIhIsar 504
QY 7753 -----GATGAATTAATAATTGAAAGACCTAAGAGATTAATTAATCTCTAAAT 7802
Db 504 gglYhISgluIAspTyrLysGlnIhISleuLeuGlyAlaSerIleTyIoluhlaGlnle 524
QY 7803 GGTTCGTCGAACTGCACGAAATAGAGAAATAGACATTAATTAATACTCACTCAGAA 7862
Db 524 uleuLysTyIySgIuLysAsp----- 533
QY 7863 CGGTCGCGAATGATCCATATGATGATCTAATAATAGTTGCGTGCATGCGTGA 7922
Db 534 -----GluAsnAlaLeuCysSerIleIleGlnAsnSerTyAlaAspLeuAlas 550
QY 7923 CAAAGTAGAGAAAGATATGTAGCAATGTGGTACTTACTCCCTAGAAATTA 7982
Db 550 pIleIleLysGlySerAspIleIleLys---AspTyTyIyGlyLysLysMetGluLys 559
QY 7983 ATTATATAGGTTTGGATACATATATGAAATGAGAAATTAATAAGGTGAA 8042
Db 569 uleuAsnLysVal-----AsnLysAspLysLysArgas 580
QY 8043 TAAATACAGATGTCAAAAGCTTCTGCTGCTGGGATGATTAATGAAAGATAT 8102
Db 580 nglu---GluSerLeuLysIlePheArgGluLysTTPASPGLuSnLysGluSnVa 599
QY 8103 TTGGAAGCAATG-----ACGTGCAAGACACACAGAGATGC 8138
Db 599 lTrrpLysValMetSerAlaValLeuLysAsnLysGluThCysLys-----AspTy 616
QY 8139 AAACTTTTAGAAAAGAAAGATGATGATTTGAACGATTAACATTAACAGATTA 8198
Db 616 rAspLysPheGlnLys----- 621
QY 8199 GTGTGACATAGAGAGATCCACTGTGATGATTAATACCTCAACGCTTTCAGAT 8258
Db 622 -----IleProGlnPheLeuArgTrrp 639
QY 8259 GACTGAATGCTGAATTAATTAATGAAAGCATGAGAAATGGAATAATTTAAAA 8318
Db 629 elYsGluTrrpGlyAspAspPheCysGluLysArgLysIleTySerPheGlnse 649
QY 8319 ATCATGATCACTGTAATAACATCTGACAGATCAAGATTAATGATTAATAAGT 8378
Db 649 rPheLysValGluCysLys-----LysLysAspCysAspGluSn----- 662
QY 8379 TGAACAGTGAACAGATGTCAGAAATTAATAATTTGTTCTTAATGAAATCTCT 8438
Db 663 -----ThrCysLysAsnLysCysSerGluTyLysTrrpIleAspLeuLysLysSerG 681
QY 8439 ATTGATTAATCAATCAATTAATTAACAAAGATTTGATGAACAACATATATACAAAT 8498
Db 681 uTyrgLysGlnValAspLysTyThrLysAspLysAsnLysLysMetTyIAspSn 701
QY 8499 CTCTACTTATGATCATGTCAAATTTTGTACAAAGTTGAAACTTTTAAAGTGAATG 8558
Db 701 e-----AspGluValLysAsn-----LysGluAl 709
QY 8559 TTCTGTGAGAGCTTTCTGCAATATCTTCATGAACAAGTAAAG---TGTGTAATTA 8615
Db 709 aAsnVal-----TyIleuLysGluLysSerLysGluCysLysAspValas 724
QY 8616 ATTAAATGAAGATGATGTTCTTCATATAGCAATATGCTTTCGAGAAACACCAAA 8675
Db 724 nPheAspAspLys-----IlePheAsnGluSerProas 735
QY 8676 AACTTAATGAAGATGCTGACGTGTACACTACTCTTAAGAAATCCATGGATTAATGTC 8735
Db 735 ngluTyrgLysAspMetCysLysLysCysAspGluIleLysTy---LeuAsnGluIleLys 754
QY 8736 TACCGATCAAAACAAAGATGATGATGAAGCAATTAACAAGCTTTT-----ACCTTCTGCTC 8789
Db 754 sTyrgProLysThyLysIhISAspIleTyArgLysAspIleThrPheSerAspThrPheGlnas 774

QY 8790 GAAGATGATTAATGATTAATTAATCTGATTAATGGAACGATACCTTGTCTTAATAGTTC 8849
Db 774 pglYthrrProIleSerIleAsnAlasn-----IleAsnGluG 787
QY 8850 AGATGATTAACAAAGGTGATATGATTCCTCCAGAGAGACATTAATGACAGACCTAT 8909
Db 787 nglInserGlyLysAspThr-----Se 794
QY 8910 CACTGCATTAATTAATGAAGAGTGTAAAGAAATTTTAAAAAACTTCTACTTC 8969
Db 794 rAsnThrGlyAsnSerGluThrSerAspSerProValSerHisGluProGluSerAspAl 814
QY 8970 TGCTTCACGAAAGCAAGCAATGTTAGGTCAAATTAATTAATGCGAAGAGAGTGTGCTT 9029
Db 814 aAlaIleAsnValGluLysLeuSerGlyAspGluSerSerSerGluThrArgGlyIlele 834
QY 9030 TGAAGCA-----AT 9038
Db 834 uAspIleAsnAspProSerValThrAsnAsnValAsnGluValIhISAspAlaSerAsnTh 854
QY 9039 GAATATATGTTATGACAGATTAATCCGATTAATTAAGCA-----ACGTATATGAT 9089
Db 854 rglInGlySerValSerAsnThrSerAspIleThrAsnGlyhISerGluSerLeuAs 874
QY 9090 GGAACCTTCATTAATCTGAAAAATTAATAATTAATTTGAACATCAATGACAAACCA 9149
Db 874 nArgThrThAsnAlaGlnAspIle---LysIleGlyArgSerGlyAsnGluInserAs 893
QY 9150 AAATCGTAAACATGCTGGGAAATTAATAGACGTCAATATGGCACGCTTGTATGTGG 9209
Db 893 rAsnGln-----GluAsnSerSer-----HisSerAspAsnse 905
QY 9210 ATATTAATGCTACTCAAAAGATTAACATTAAGATTAAGAA----- 9246
Db 905 rglYSerLeuThrIleGlyGlnValProSerGluAspAsnThrGlnAsnThTyIAspse 925
QY 9247 -----GAGTGTGTCATTAACCAAGAGATGAAGAAAC 9278
Db 925 rGlnAsnProhISArgAspThrProAsnAlaLeuAlaSerLeuProSerAspLysI 945
QY 9279 TAACTAGTTTCTGCTGTTGTTAATGAATGGCAAGCAAGCATGTAAGAAAGAAACA 9338
Db 945 eAsnGlu-----IleGluGlyPheAspSerArgAspSerGluSnG 950
QY 9339 TGTAACTGATTCATTAATAACAAATGCTCCTCAACAGAGATTAATTTGAAGGTC 9398
Db 960 yArgGlyAspThr-----ThrSerAsnThhISAspValArgArgTh 974
QY 9399 AGAATTAATTAAGACAACTGGATGTCAGAAATGATTAATTAATTAATAGCTTGAATAT 9458
Db 974 rAsnIleValSerGlu-----ArgArgValAsnSerhISas 986
QY 9459 ATTGATTAATAATCAATGCAATGCAATTAATTAATTAATTAATTAATTAATTAAT 9507
Db 986 rPheIleArgAsnGlyMetAlaAsnAsnAsnAlahIShISGlnTyIleThrGlnIleG 1006
QY 9508 -----AAAGTCAATCTTCAGTAATATGACAAAT----- 9537
Db 1006 uAsnAsnGlyIleIleArgGlyGlnGlnGlnSerAlaGlySerValAsnTyIysAs 1026
QY 9538 -----AAACATCTGAAGAAATGTCATGATCA 9566
Db 1026 rAsnProLysArgSerAsnPheSerSerGluAsnAspHisLysLysAsnIleGlnGlu 1046
QY 9567 TATTAATCAAAAGATTCATAGCCGCTTGGAGTTAAATGATTAATGAATTAATATAC 9626
Db 1046 r---AsnSerValAspThrLysArgValArgGlu-----GluIleLys 1060
QY 9627 AGGACAAAAAATTAATGAA---AATATGAAATTC 9657
Db 1060 sLeuSerLysGlnAsnLysCysAsnSnGluTy 1071

D	b	55	glutrylsglnsnshnvalnalspillellephegilythleuansntylrgluty-	72
O	y	7513	ATTGAGCTCTCAAGAGAAAAGAAAGTAAATAGTGGCATATATAATATCTTGGCAT	7572
D	b	73	AsnasphecylysylsglulysProgluLeuvalSerlaalaIalysylrshn	89
O	y	7573	CTTAGAAACCTTATCGACCTGATTAATATATAGGAAGAAACCTTGTAATAGA	7633
D	b	90	LeulysalaleproasnlaIalysSerProArgrilleltyrlysserlys	107
O	y	7633	GAAGAAATCGTTTT-----AAGCTAGATTATGGAAGAAATGT	7671
D	b	108	glulserSerValphegilyCyslysthrlyslleSerlyvalylslysltyrshpncys	127
O	y	7672	TACAAAATTCAAGCTCTATCAGAGAAAGAAAGAGATATGTCTACCTCAAGAAAGAA	7731
D	b	128	TySerAsnshnlys---ValIthrltyrProgluIvalIcysglYlProProArgrlncin	146
O	y	7732	CATATGTGCTTAAGAAATTTAGATGAATTAATA-----ATTGAAGACTTAAAGAT	7782
D	b	147	gluleucysleuglYtyrllepheleuilelrgaspolyasnoluglulyleuylasp	166
O	y	7783	AGTAATATCTCTCAAAATGGTTCGTCGACCTGCACGAAATGAGAAATACACATATA	7842
D	b	167	Hls-----leasnlysalalaasnlyrglulnlaIethlshleuyls	180
O	y	7843	AAAACTTCACTCAGAGAACGGGTGGCAATGCAATTCATATATGATCTATGAAATAT	7902
D	b	181	glulysTyrglulasnalaIglYl-----AspyslilecysasnlaIalleuugly	197
O	y	7903	AGTTTCGCTGATCTGGGTGCACATAGTTAGAGAACAGATATGTTCAGATTGGTGGTTAC	7962
D	b	198	SerTyrlalaspIleglYaspllevalalrglYleuaspValItrparg-----	213
O	y	7963	TTACCTCCCGTAGAATA-----AAATPTATAAGCTTTTGGATATGATATATGCA	8013
D	b	214	-----AspIleasnIthAsnlytleuSerGilylshPheglulysllepheet	229
O	y	8014	AAATGAGAAATTAATAAAGTACAGATTAATATACACAGATGTCAAACTTTGCTTCT	8073
D	b	230	glylglYlgnanSeratrglylsglnasnaspasnnglu-----Argasn	245
O	y	8074	GCTTGCTGGATGCTATATGAAAAGATATTTGGAAAGCAAT-----	8115
D	b	246	IystrIprglulysglnalghasnleuilelItrpSerSermevalylshsileprolyls	265
O	y	8116	-----ACGTCAAAAGCACAGAGATGCAAACTTTTACAAAAGCAGATGATGAGA	8169
D	b	266	glylstrIthrcylslyl-----ArgHlsAsnshn	274
O	y	8170	TTTGAAGCCATACATTAAATACAGAGATAGGTGACATTAAGACGATCCACCTGTTGAT	8229
D	b	275	Pheglulys-----	277
O	y	8230	GATTATATACCTCAACGGTTTCGATGATGACTGATAGTGTGGAATATATGTAAGCA	8289
D	b	278	-----lleproglInpheleuartrgrpleuylsglultrpglYasplunphecysglulgu	295
O	y	8290	CTGATGGAAGCAATTTGAAAATTTAAAAATCATGTGATCTCACTAAAAACATTCGACGA	8349
D	b	296	MetelYlthrcIuVallylsglnleuglulysIleCysgluasnIlysaencySerGilyls	315
O	y	8350	TGCAAGATGATTATGATGAATAATAGTGCAGACAGTGTAAACGAGATGTCAAGAAATAT	8409
D	b	316	-----lyscylsAsnlaIcysSerSertyr	324
O	y	8410	AAAAATTTGTCTTAATAGGAATCTCATATGCATATACATCAATAATACAAAGAA	8469
D	b	325	glulysTrIlelylsglnulrglylsasnIulYrIasnleuGlnSerlylsPheaspser	344
O	y	8470	TTGATGAACAACCATATATATACAAAAATCTCTCATATGATCATGTGCCAAATTTTGTA	8529
D	b	345	-----AspYlshlyleuasnulyslylsasnshnleuYlryshn-----	356

```

OY 8530 CAAAGCTGAAACCTTTTAAAGTGAAGTCTGTTGAGAGCTTTTCTGAATATCTTCAT 8589
DB 357 -----LysPheGluAspSerLysAlaTyrLeuArg 366
OY 8590 GAAACAGTAG--TGTGAATATTAATTAAGAAAGATGCGTCTCCATATA 8646
DB 367 SerGluSerLysGlnCysSerAsn1LeuPheAsnAsp----- 379
OY 8647 CGAACATATGCTTTCGAAACACCAAAAGATTAAGAAGCTTCAGT---TGTACA 8703
DB 380 GluThrPheThrPhe-----ProAsnLysTyrLysGlnAlaLeuGlySerMetAlaCysGlu 396
OY 8704 CTACCTCTGAAGAT-----CCATTGGATAT--TGTCTACCGATCA 8745
DB 397 AsnProSerSerSerLysAlaLeuLysPro1LeuLysThrAsnValPhePro1LeuGln 416
OY 8746 AACAAAGATGATGTAAGAAATTAACAACCTTCTGCTCCGAGAAATGATTATGAT 8805
DB 417 SerLysLys-----SerGluLeuSerSerLeuThrAspLysSerLysAsn----- 431
OY 8806 AATTAATCTGATTAATGGAACCATCTGTTCTTAATAGTTCAGATGATAACAAAGCT 8865
DB 432 -----ThrProAsnSerSerGlyGly----- 439
OY 8866 GTATTGATTCCTCCAAAGAAAGACATTTATGTACAAAGCTATACGATATTAATAT 8925
DB 440 -----AsnTyr 441
OY 8926 AGAAAGTGATTAAGAAATTTTAAAAA-----NAAGCT 8961
DB 442 -----GlyAspArgGln1LeuSerLysArgAspAspValHisHisAspGlyProLysGlu 459
OY 8962 CTACTTCTGCTTTCAGTCAGACACAATTTAGTCAAAATATTAATCGAAGACAG 9021
DB 460 ValLysSerGlyLysGluValProlLys1LeuAspAla1ValLysThrGluAsnGlu 479
OY 9022 TTGTGC-----TTGAGCGAATGAATATAGTATGACAGTATTCGAT 9066
DB 480 PheThrSerAsnArgAsnAsp1LeuGlnLysGluLysSerLysGlyAspHisSerSer 499
OY 9067 ATAATTAAGAACTGATGATGACACTTCATTATCGAAAAAATTAATAAAATATTT 9126
DB 500 ProValHisSerLysAsp1Le-----LysAsnGlu 509
OY 9127 GAACATCAAAATGAACCAACCGAAATCTTAAACATGCTGGGAAATATAGACCTCAG 9186
DB 510 GluProGlnArgValValSerGluAsn----- 518
OY 9187 ATATGCGACGCTATGTTATGTGATTAATAATGCTACTTCAAAAGTAACATTAGATGAA 9246
DB 518 ----- 518
OY 9247 GGTGTGTCATTAACCAAGATGAAGAAACATACATCTTCTTGCTGTTAATTGAA 9306
DB 519 -----LeuProLys1LeGlu----- 523
OY 9307 TGGGCAACCAAGCATTAAGGAAAGAAACATGTAAGTATTCATTAACCAAAATGCT 9366
DB 524 -----GluLysMetGluSerSerAspSer1LeuPro1Le----- 534
OY 9367 CCTCGTTCAAGAGATATTTTGAAGCGTCGAATTTTAAGACACCTGGATGTCAG 9426
DB 535 -----ThrHis1LeuGlnAlaGlu----- 540
OY 9427 AATGATATTAGAAAATATATTAGCTTGAAATATTGATTAATAAAATACATGGAATCTA 9486
DB 540 ----- 540
OY 9487 AATATATAATTAAGCAATTAAGATCAATCTCAGTAATATAGACATTAACCATCT 9546
DB 541 -----LysGlyLysSerSerAsnSerSerSerAspAsnAspProAla 553

```

```

OY 9547 -----GAGAA 9552
DB 554 ValValSerGlyArgGluSerLysAspValAsnLeuHisThrSerGluArgLeuLysGlu 573
OY 9553 AATGTCAGCATATATTAATAATCAAAAGATTCATGCGCTTGGAGTTA----- 9603
DB 574 AsnGluGluGlyVal1LeuLysThrAspAspSerSerLysSer1LeuLysSerLys1Le 593
OY 9604 -----AATGATATAATGAATGATGTTACAGCAACAAAATAAT 9642
DB 594 ProSerAspGlnAsnAsnHisSerAspLeuSerGlnAsnAlaAsnLysAspSerAsnGln 613
OY 9643 GAAATATATGA-----TTCAAAGAGTACTAAAAAATTAATATCT 9684
DB 614 GlyAsnLysGluThrLeuAsnPropSerThrGluLysAsnLeuLysGluLeuHis 632
OY 9685 GCTTTATATTGTTGTAAGATGAACACCAAAATCATGTAATGATGAAATATATAA 9744
DB 633 -----TyrLysThrSerAspSerAspAspHisGlySer 643
OY 9745 GAAGAGACAAACAGTTCGCTCAAGCAGCTGATTTCTTACACCCCATGTAGATTC 9804
DB 644 Lys1LeuLysSerGlu1LeuGluProLysGluLeuThrGluGluSerProLeuThrAspLys 663
OY 9805 TTCTATCAAGACCTTTATCTCAACACATCGAGTACACAAATATGCTTCAAAATGAT 9864
DB 664 -----LysThrGluSerAla1Ala1LeuLysAspLysAsnHisGlu 676
OY 9865 ATATGAAGATGCT----- 9879
DB 677 SerValLysSerLeuAsp1LeuPheGlnSerGlu1LeuHisAsnSerAspAsnArgAspArg 996
OY 9880 ---ATCTGCTGTTATTTGTA-----TCGGCGTTCAGTTTGAATAGCGCTT 9921
DB 697 1LeuValSerGluSerValValGlnAspSerSerSerLysSerSerMetSerThrGluSer1Le 716
OY 9922 CATTTATGAAGAAAATTAATCAATGCTGTCGACTGTTGCGATGACTGAAT 9975
DB 717 ArgThrAspAsnLysAspPheLysThrSerGluAsp1LeuAlaProSer1LeuAsn 734

RESULT 10
US-09-924-154-16
; Sequence 16, Application US/09924154
; Patent No. US20020127241A1
; GENERAL INFORMATION:
; APPLICANT: Natum, David L.
; TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
; FILE REFERENCE: 05213-0465 43170-262105
; CURRENT APPLICATION NUMBER: US/09/924/154
; PRIOR APPLICATION NUMBER: US 60/223,525
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 972
; TYPE: PRF
; ORGANISM: Mammalian
US-09-924-154-16

Alignment Scores:
Pred. No.: 1,81e-16 Length: 972
Score: 382.50 Matches: 253
Percent Similarity: 31.86% Conservative: 157
Best Local Similarity: 19.66% Mismatches: 450
Query Match: 1.97% Indels: 49
DB: 10 Gaps: 43

US-10-087-013-1 (1-10628) x US-09-924-154-16 (1-972)
OY 3151 TTAAGTGTATGACCGAATGGCAGAAATGCTACTGCAAGCTGAGAAAAAGAGTATGAT 3210
DB: 3151 TTAAGTGTATGACCGAATGGCAGAAATGCTACTGCAAGCTGAGAAAAAGAGTATGAT 3210

```

[illegible]

230 MetTrnGluTrpAlaGluTrpTyrCysLysAlaGlnSerGlnGluTyrAspLysLeuLys 249

Query match: 1
B: 9

Read, No.:	5,9e+14	Length:	1115
Score:	343.50	Matches:	243
Percent Similarity:	32.46%	Conservative:	177
Best Local Similarity:	18.78%	Mismatches:	435
Query Match:	1.77%	Indels:	439
B:	9	Gaps:	60

QY 8917 TATATATGAAAAGCTGATTAAGAAATTTTAAAAAAACCTGCTACTGCTGCTTC 8976
DB 601 GlysAlaValAsnSerGlyGlnAspSerThrThrGlyAlaValThr----- 617
QY 8977 AGTCAAGACAAATTTGAGTCAAAAATATTAATCGAAGAAAGAGTTCCTTGAGCCA 9036
DB 618 GlyAspGlyGln--AsnGlyAsnGlnThrProAlaGlnSerAspVal----- 632
QY 9037 ATGAAATATATGATTATGACGATTATTCGATTAATTAAGAAAGTATATGATGACACT 9096
DB 633 -----GlnArgSerAspIleAlaGlnSerValSerAlaVal----- 646
QY 9097 TCATATTCGAAAAAATTAATAATATTTGAACAATCAAAATGACCAAGAAATCGT 9156
DB 647 AspProGlnValSerValSerValSerValSerValSerValSerValSerVal 654
QY 9157 AAAACATGCTGGGAAATATATAGACGTCAGATATGCGACGCTATGTATGATATAA 9216
DB 654 ----- 654
QY 9217 ATTGCTACTTCAAAAGTAAACATTAGATGAGAGTGTGTCAATTAACAAAGATGAGAA 9276
DB 655 -----ArgSerAspAsp----- 658
QY 9277 ACTAATCAGTTTCTGCTGTTGTTAATGGAATGGCAAGCAAGCATGTRAGAAAGAAA 9336
DB 659 ThrAlaSer-----ValThrGlyIleAlaGlnAlaGlyLysGlnLeu 673
QY 9337 CATGTAGTCAATTAATAAACAATGCTCCTGTCACAAAGCAATTAATTTGACGC 9396
DB 674 GlyAlaSerAsnSer-----ArgProSerGlnSerThrValGlnAla 687
QY 9397 TCAGATATATTAAGAACCTGATGTCAGATGATATTAAGAAATATATGCTGAAT 9456
DB 688 Asn-----SerProGly-----AspAspThrValAsnSerAlaSerIlePro 701
QY 9457 ATATGATATAAATAATACATGGAAT--CTAATATATAATATTAACCATTAAGAAAT 9513
DB 702 ValVal-----SerGlyGlnAsnProIleuValThrProTyrAsnGlyLeuThrHis 718
QY 9514 CAATCTTACGTAATATATGACAAATTAACCATCTGAGAA-----AATGTCAGCTATAT 9567
DB 719 SerLysAspAsnSerAspSerAspGlyProAlaGlnSerMetAlaAsnProAspSerAsn 738
QY 9568 ATAAATCAAAAAGATTCTCAATGCGCTTTGGAGTTAATGATATAATGAATGTTACA 9627
DB 739 SerLysGlyGlnThrGlyLys-----GlyGlnAspAsnAspMetAlaLys 753
QY 9628 GGAACAAAAAATTAATGAATAATGAATTAAGAAAGTACTAAAAAATATATCTCGT 9687
DB 754 AlaThrLysAspSerSerAsnSerSer----- 762
QY 9688 TTATATTTTGTGAAGATGAACAACACAAAAATCATCTACTAGATGGAATATTAAGAA 9747
DB 763 -----AspGlyThrSerSerAla 768
QY 9748 GAAGAGCAAAAGTTCGCTCTAAAGCACTATTTCTTACACCCATGGATTTCTTC 9807
DB 769 ThrGlyAspThr-----ThrAspAlaValAsp----- 777
QY 9808 TATCAACACCTTATATCTCAACACATCGAGTACACAAATATGATCTTAAATGATATA 9867
DB 777 ----- 777
QY 9868 TTGAAAGTAGTATCTCTGTTTATGTATCGCGCTTAGCTTGAATAGCGCTTCAATTC 9927
DB 777 ----- 777
QY 9928 ATGAGAAAAAATTAATCAATCTGTGTGGACTTGTGATATCTGATATCCCGAAGA 9987
DB 778 -----ArgGlnIleAsn----- 781

QY 9988 GAGTATGGAATGCTTACCTGATGATCCAAAAATAGCTACATACATTAATGAGAGCTCCA 10047
DB 782 ---LysGlyValAlaPro-----GlnAspArgAspLysThrValGlySerLysAspGly 798
QY 10048 TATAAGCAAAACATATATATATATGAGAGAGTACTACTAGGAGTACAGAT----- 10101
DB 799 GlyLys-----AspAsnSerAlaAsnLysAspAla 809
QY 10102 -----AAATATATGAGGACTTAACTTCTCTGATTAATTAATTAATTAATTAAT 10146
DB 810 ThrValValGlyGlnAspArgIleArgGlnAsnSerAlaGlyGlySerThrAsnAspArg 829
QY 10147 AGTGAATGGAAGATGGAATTAATATATATATATGACAGTACTCTTAATTAATAA 10206
DB 830 SerLys---AsnAspThrGlyLysAsnGlyAlaSerThrProAspSerLysGlnSerGlu 848
QY 10207 ACATGATGAGAGTATGATCTAGTAACCAATCAAAAAGGATATACCAAGTGAATGACCA 10266
DB 849 AspAlaThrAlaLeuSerLysThrGlnSerLeuGlnSerThrGlnSerGlyAspArgThr 868
QY 10267 AGTATGATACACACGTCAG-----AATGATTTATGATGATGATGATGATGATGAT 10320
DB 869 ThrAsnAspThrThrAsnSerLeuGlnAsnLysAsnGlyLysGlnLysAspLeuGln 888
QY 10321 AAACATGATTTTGTATCTCAATATATTAATTAATTAATTAATTAATTAATTAATTAAT 10380
DB 889 LysHisAspPheLysSerAsnAspThrProAsnGlnLysProAsnSerAsp---GlnThr 907
QY 10381 GCATATTTCCATATGATACAGAACCTTAATAGCTTATATATCTGATATATCTGAGAAAA 10440
DB 908 ThrAspAlaGlu---GlyHisAspArgAspSerIleLysAsnAspLysAlaGlnValArg 926
QY 10441 CCTTTTATATCTATATCTATGATAGGATTTATATATCTGGAAGAAATTAATTAAT 10500
DB 927 LysHisMet-----AsnLysAspThrPheThrLysAsnThrAsnSerHisHis 942
QY 10501 ATTAATATGAGTACTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10560
DB 943 LeuAsnSerAsnAsnAsnLeuSerAsnGly---LysLeuAspIleLysGlnLysLysIle 961
QY 10561 AGAGGTATATATTAATTAATGATTCATCTAGTGTGCTTAAC 10602
DB 962 ArgAspValLysAlaThrArgGlnAspIleIleLeuMetSer 975

RESULT 13
US-09-815-242-12996
; Sequence 12996, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIORITY FILING DATE: 2001-03-21
; PRIORITY FILING DATE: 60/191,078
; PRIORITY FILING DATE: 2000-03-21
; PRIORITY FILING DATE: 60/206,848
; PRIORITY FILING DATE: 2000-05-23
; PRIORITY FILING DATE: 60/207,727
; PRIORITY FILING DATE: 2000-05-26
; PRIORITY FILING DATE: 60/242,578
; PRIORITY FILING DATE: 2000-10-23
; PRIORITY FILING DATE: 60/253,625
; PRIORITY FILING DATE: 2000-11-27
; PRIORITY APPLICATION NUMBER: 60/257,931

Db 2450 AlaTyrAsnAlaAlaLysAsn-----IleVal 2458
OY 1738 AATGATTAATCTTAAGTATATATATCATTAATTTTGAATATATGGCTTACATATTTA 1797
Db 2459 AsnGlySerProAsnValIleThr-----AsnAlaAlaAspValThrAlaAla----- 2474
OY 1798 TTAAAGGATCTATTAAAGTGAATGACAACCTTAACCTGTATTAATTAATACACACG 1857
Db 2475 -----ThrGlnArgValAsnAsnAlaGlnThr 2483
OY 1858 CATTTATGATGAATGTACAGAAATTCCTTATGTTTGACAGATGGCTTAACAAAA 1917
Db 2484 GlyLeuAsnGlyAspThrAsn-----LeuAlaThrAlaLys 2495
OY 1918 GAGAGAAATGAAATAGTATTAAGAAATCTTCACAAAAAAAGAAATATACAGCAATCG 1977
Db 2496 GlnGlnAlaLysAspAlaLeuArgGlnMetThrHisLeuSerAspAlaGlnLysGlnSer 2515
1978 TATATATGTAATTAATTAATCTTTTGAAGCTTATTTTAAAGTTATGATTAACCTT 2037
2516 IleThrGlnLysIleAspSerAlaThrGln----- 2525
OY 2038 GACAAGATGATGAACAAATGGAAGAACTTAATGGAATAATTAAGAAAAAATATGAG 2097
Db 2526 -----ValThrGlyValGlnSerValLysAspAsn 2535
OY 2098 TTTTCCAAATTTGCAAAATTAATAGGACTATTTAGAGAAATGCAATAGAACTCTTGTAGAT 2157
Db 2536 AlaThrAsnLeuAspAsnAlaMetAsnGlnLeuArgAsnSerIle---AlaAsnLysAsp 2554
OY 2158 CACTTAATAAGAACT-----GCCAGATATGTAAACAACTAATACAAAC 2202
Db 2555 AspValLysAlaSerGlnProTyrValAspAlaAspArgAspLysGlnAsnLalaTyrAsn 2574
OY 2203 GAAGCA-----TGTGAACATCCCATATATGCAACAAACCCGTGTGTAAACCT 2253
Db 2575 ThrAlaValAlaThrAsnAlaGlnLysIleIleAsnAlaThrSerGlnProThrLeuAspPro 2594
OY 2254 CGTGAAGAGCAGCAACCACTTAATAATTAAGAAATATGCAATCTTTAAAGAGCT 2313
Db 2595 SerAlaValAlaThrGlnAlaAlaAsnGlnValSerThr-----AsnLysThr 2609
OY 2314 GCATACGAGAGAACGCAAAATCGTGCTTCATTAATTTGAAGAAAGCAGCAAGAGT 2373
Db 2610 AlaLeuAsnGlyAlaGlnAsnLeuAlaAsnLysLysGlnGlnThrThrAlaAsnIleAsn 2629
OY 2374 ATATATAACGCTGGGGGTGAGAAAGCACTCAAGCAATTTATGTAGATAATATGATA 2433
2630 -----GlnLeu 2631
OY 2434 AATCATTTCTAATCGT-----AATCTTGGTTTTTCAATGGACCATGT 2475
Db 2632 SerHisLeuAsnAsnAlaGlnLysGlnAspLeuAsnThrGlnValAlaThrAsnAlaPro--- 2650
OY 2476 GATGGCAAAAGCAGAGTGATGTACAAACA----- 2508,
Db 2651 ---AsnIleSerThrValAsnGlnValAlaLysThrLysAlaGlnLysAlaMet 2659
OY 2509 ---AGATTTCGTAGAACTGAATGGAAGTCCGGGAACACATCCGTAAGAGTAC 2655
Db 2670 GlnArgLeuIleAsnGlyIleGln-----AspLysAspGlnValLysGlnSerVal 2686
OY 2566 GAAAGTGTATTATTCCTCTAGAAAGACATATATGATACCAATTTGAGACATTTA 2625
Db 2687 AsnPheThrAspAlaAspProGlnLysGlnThrAlaTyrAsnAsnAlaValThrAlaAla 2706
OY 2626 CAATGATGATGATCAACCCCTTAATGTAATATTTGTATGATTTAGTTAATATTCCTTT 2685
Db 2707 GlnAsnIleIleAsnGlnAlaAsnGlyThrAsnAlaAsnGlnSerGlnValAlaAla 2726
OY 2686 TTGGGGGATGTTCTTCTATCAACAAATATGAAACAAAGAAATATACGAATGTATAA 2745
2686
Db 2727 LeuSerThrValThrThr----- 2733
OY 2746 GAAAGAAATTAACCTAAAGGCCCAAGAGTAACTAGTACCCAAACACACAGCA----- 2799
Db 2734 ---LysGlnAlaLeuAsnGlnLysAspArgGlyValThrAspAlaLysAsnAlaAsnGln 2752
OY 2800 ACTATCTGTCGACCTATACGTTACAGTTTTCAGAT-----ATAGCTGATATA 2847
Db 2753 ThrLeuSerThrLeuAspAsnLeuAsnAsnAlaGlnLysGlnValAlaValThrGlnLysAsnIle 2772
OY 2848 ATTCAGAGAAAGATCTCTGGAGAAAGAAACGGTGACATGTGTAACCTCAAGACATTTG 2907
Db 2773 AsnGlnAlaHisThrValAlaGlnValThr---GlnAlaIleGlnThrAlaGlnLysLeu 2791
OY 2908 GAACTGTTTTTGGTAATATACATATGATCTACATAGGCAAA-----GGAAT 2955
Db 2792 AsnThrAlaMetGlyAsnLeuLysAsnSerLeuAsnAspLysAspThrThrLeuGlySer 2811
OY 2956 GATTAATATATGATGATGCCCCCAATATTTAAATTTGACGGAAATTCGTGGGAA--- 3012
Db 2812 GlnAsnPheAlaAspAlaAspProGln-----LysLysAsnAlaTyrAsnGln 2827
OY 3012 ----- 3012
Db 2828 AlaValHisAsnAlaGlnAlaAsnIleLeuAsnLysSerThrGlyThrAsnValProLysAsp 2847
OY 3013 -----GCTAATAGACCAAGTATGGAGAACCATGAATGTGATATATAA 3057
Db 2848 GlnValGlnAlaAlaMetAsnGlnValAlaAsnAlaThrLysAlaAlaLeuAsnGlyThrGln 2867
OY 3058 TATTTGAAGATTAATATCGGAGCAACCAATCAACACAAAGTATATTCGGATATAGAT 3117
Db 2868 AsnLeuGlnLysAlaLysGlnHisAlaAsnThr-----AlaIleAspGlyLeuSerHis 2885
OY 3118 CATACACCATTTGGAT---GATTATATCCACACAAATTAATAGATGATGACGAATGGCA 3174
Db 2886 LeuThrAsnAlaGlnLysGlnAlaLeuLysGlnLeuValGlnGlnSerThrThrValAla 2905
OY 3175 GAATGTACTGCAAGTGTGCAAAAGAGAG-----TATGATAGTTGAAG 3219
Db 2906 GlnAlaGlnGlnAsnGlnGlnLysAlaAsnAsnValAspAlaAlaMetAspLysLeuArg 2925
OY 3220 GAGAACTGTAAAGCTGTAAAGATTAATGCTCAAGGCTGTCAAGAGAGTGTGT 3279
Db 2926 GlnSerIleAlaAsp-----AsnAlaThrThrLysGlnAsn--- 2937
OY 3280 ACAGTTGTACGAAGTGCACAGAAAGCTTGAATATATATATATTAATAGAAATTATGG 3339
2938 -----GlnAsnTyrThrAspAlaSerGlnAsnLys 2947
OY 3340 AAGAACATGGAATATATATCATGATTAATACAAAGATTAATCATGACACAGCA----- 3393
Db 2948 LysAspAlaTyrAsnAsnAlaValThrThrAlaGlnGlyIleIleAspIleThrThrSer 2967
OY 3394 -----CAATGTCTGTACTAATTAATGATGTAAGCTTCAGTACAGTCCCAAAAT 3444
Db 2968 ProThrLeuAspProThrValIleAsnGlnAlaAlaGlnValSerThrThrLysAsn 2987
OY 3445 CATATGACAGAGAAATTTATGAAATTTTGGGAATTAATCCAAACAAATGGTGCAAA 3504
Db 2988 AlaLeuAsnGlyAsn-----GlnAsnLeuGlnAlaAlaLysGlnGln---Ala 3002
OY 3505 AGTAATTAAGAGTGTACTAGATGAAGTGTCTGATGTACTATACACACAGATGAA 3564
Db 3003 SerGlnSerLeuGlySerLeuAspAsnLeuAsnAsnAlaGlnLysGlnThrValThrAsp 3022
OY 3565 AATGTT---GGAGCAATTCATCATGATACAGAAATTTTGAATGTGTGCTCACTCAAAAT 3621
Db 3023 GlnIleAsnGlyAlaHisThrValAspAlaAlaAsnGlnIleLysGlnAsnAlaGlnAsn 3042
OY 3622 GAGTTTGTATGAAGAAAGATGATGTAAGATTAACGAATATATGCTTATAGATATAA 3661
3043 -----LeuAsnThrAlaMetGlyAsnLeuLysGlnAlaIleAlaAspLys 3057

QY 3682 CCACAGGACCATGATGTCGCTGCTGTTGTAAGTGCATCCAAACCGCAAGGTCAG 3741
Db 3058 -----AspAlaThrIysAlaThr 3063
QY 3742 ATAAAAACGAAAAAAACCGAAGAAAGATACGGAATGTAAAAACATG----- 3792
Db 3064 ValAsnPheThrAspAlaAspIleAlaIysGlnIleAlaThrAsnThrAlaValThrAsn 3083
QY 3793 -----AATGATATACTTAAAGAAAAAGATGGAAG----- 3822
Db 3084 AlaGlnAlaIleIleSerIysAlaAsnGlyGlyAsnAlaThrGlnAlaGluValGlnGln 3103
QY 3823 -----AAACAGTGAAGACATGTCATCCAAAAAGATAGTAATGATATCCGATGG 3876
Db 3104 AlaIleIysGlnValAsnAlaAlaIysGlnAlaLeuAsnGlyAsn----- 3118
QY 3877 CAATGGCAATATATAATTAGTCGAGACCTCGTGTGTATGCCCCCTAGAACAA 3936
3119 -----AlaAsnValGlnHisAlaIysAspGluAlaThrAlaLeu----- 3131
QY 3937 AAGTTATGCGTACATTTCTGGCAATGATTAATGAATAAAAATTCACATCACAATG 3996
3132 -----IleAsnSerSerAsnAspLeuAsnGlnAlaGlnIlyAspAla 3145
QY 3997 AATTTAAAAGACCTTTCATCAATCTGCAGACGACAAACATTTCTCATGTATTTAT 4056
3146 -----LeuIysGln-----GlnValGlnAsnAlaThrValAlaGlyValAsnAsn 3161
QY 4057 TATTAAGTGAAGATGTCGAGCAAGAAATGACGATTAAGATTAATAAGCAAAAT 4116
3162 ValIysGlnThrAlaGlnIleuAsnAsnAlaMetThrGlnLeuIysGln----- 3178
QY 4117 CCTCCGCAATTTTGTAGATCATGTTCTACATTTGGATTAATGAGATTTTATTT 4176
3178 ----- 3178
QY 4177 GGAACGATATATCAAAAGGTCATGTCGAGCAAGTAAACSTAAAGACAAATGATTC 4236
3179 -----GlyIleAlaAspIysGlnGlnThrIysAlaAspGly 3190
QY 4237 CTTTCAAAATGTGACCAAAATCTCTAATGGAAGAAAAACGCCAAGATGTCGACA 4296
3191 AsnPheValAsnAlaAspProAspIysGlnAsnAlaThrAsn----- 3204
QY 4297 GAACATAGTCATGAGATATGAGACGATATGTCATGATTAATAATGGGCAAAA 4356
3205 -----GlnAlaValAlaIysAlaGluAlaLeuIleSerAlaThr 3217
QY 4357 AAAGATGATTTTACGAAACATACGCTTACACACACGTCAAATTTAGTCAAAAC 4413
3218 ProAsp-ValValIValThrProSerGluIleThrAlaAlaLeuAsnIlyValThrGlnAla 3237
QY 4414 ACCACGTTGGAGAAATTTGCCAAGACCCAGTTTTCAGATGGCTAACCGATGGTAC 4473
3237 A-----LysAsnAsp-----LeuAsnGlyAs 3244
QY 4474 GACGACTATGCTATACACGACAAATAATTTGAAGATGTCGACGAAATAATGATGTA 4533
3244 nThrAsnIleuAlaThrAlaIysGlnAsnVal-----GlnHisAlaIleAspI 3260
QY 4534 AATGACCAATTAAGTGTATACAG----- 4558
3260 nLeuProAsnIleuAsnGlnAlaIysAspGluIlySerIysGlnIleThrGlnAlaThr 3280
QY 4559 -----AATGATATAAGAAATGCGAGGATAGCTAGTAAATAT-----ATGAA 4598
3280 nLeuValProAsnValAsnAlaIleGlnIleAlaAlaThrThrLeuAsnAspAlaMetThr 3300
QY 4599 AAAAAAAGAGTGCATTCAC-----CAAGATAAATATTACAA 4637
3300 rGlnLeuIysGlnGlyIleAlaAsnIlyAsnAlaGlnIleIysGlnIlySerGluAsnThrIysAs 3320

QY 4638 GCATGACCCGCAAAAA----- 4656
Db 3320 PalaAspThrAspIysGlnThrAlaIysAspAsnAlaValThrIysAlaGluGluIleuLe 3340
QY 4657 -----AGATTGATAGACACACCATGGGTATGCTTACAGACTA 4697
Db 3340 uIysGlnThrThrAsnProThrMetAspProAsnThrIleGlnIleAlaLeuThrIysVa 3360
QY 4698 TACTGACCAAT----- 4710
Db 3360 LAsnAspThrAsnGlnAlaLeuAsnGlyAsnGlnIysLeuAlaAspAlaIysGlnAspAl 3380
QY 4711 -----GCAACGATTTACTTGAAC-----AGCAATTTACTGC 4742
3380 aIysThrThrIleGlyThrLeuAsnHisIleuAsnAspAlaGlnIlyGlnAlaLeuThrTh 3400
QY 4743 TAGTGTGTGATTAAGCCGTGGAAGTCCCTGTG-----GTACAAAGAAATATCAAT 4796
3400 rGlnValGluGlnAlaProAspIleAlaThrValAsnAsnValIysGlnAsnAlaGlnAs 3420
QY 4797 GTTAGAA-----AAACAGCTTACTATGATGCCGACAAACATTTGTG 4838
3420 nLeuAsnAsnAlaMetThrAsnLeuAsnAsnAlaLeuGlnAspIysThrGlnThrLeuAs 3440
QY 4839 GTGCACAAAATTTTATGAATATGAC-----GACAAATATACTACATTTTCAG 4886
3440 nSerIleAsnPheThrAspAlaAspGlnAlaIysIysAspAlaIyThrAsnAlaValSe 3460
QY 4887 TAAAGTAAAGTGCAAAGGATTTAGTAAGGCAACACACAGGTCATTAAGTGGCAAA 4946
3460 r-----HisAlaGluGlyIleLeuSerIysAlaAsn----- 3470
QY 4947 CAAGGTCCTAATTAACATACTTAACATTAACATTAACATTAACATTAACATTAAC 5006
3471 -----GlySerAsnAlaSerGlnThr-----GluValGlnGlnAlaMet 3483
QY 5007 TCGTCGACTACATATATGTTTTCATGCAATGCAATTTATACAGATCCAGAACTTA 5066
3483 rGlnThrValAsnGlnAlaIysGlnAlaLeuAsnGlnIlyAsnAspValGlnThrAlaIly 3503
QY 5067 AGATGAAATATGGGTTCCGAAAAAGATTAAGTAAGTGGCGCAAGGAGGATACAAAT 5126
3503 sAsp-----AlaAlaIysGlnValIleThrAsnAlaAsnAspLe 3516
QY 5127 GGCTCAATACACAAAGAA-----AAAAAGAAAAAGAAATTAACATCGGATGC 5180
3516 uAsnGlnAlaGlnIlyAspAlaLeuIysGlnIlyValAspAlaIleAlaGlnThr----- 3533
QY 5181 GCACAAATTTCTTATGAGGTCCCGCTTGTACTGCTATGAAATATATGTTTATGATTT 5240
3534 -----ValAlaAsnValAsnThrIleIysGlnThrAlaGlnAspLe 3547
QY 5241 AAGAGATATATCTGATGATGATTAATTTGGAAGATGAAGAAACAAACGACGAGAAA 5300
3547 uAsnGlnAlaMetThrGlnLeuIysGlnIlyIleAlaAspIysAspGlnThrIysAlaAs 3567
QY 5301 TTGGAAGAAATA-----TTTAACAAATAATGGAACATGCTGGCA 5342
3567 nGlyAsnPheValAsnAlaAspThrAspIysGlnAsnAlaIyThrAsnAlaValAlaIle 3587
QY 5343 AGGAAGTATAGTACTACAGGAATCCCGTACTACTGCGCGCAAAATTTTTCGACAA 5402
3587 sAlaGluGlnIleIleSerGlyThrProAsnAla-----AsnVa 3600
QY 5403 AATAAGCAATGTGTGGAACGCAATG-----ATATCGGGGTACAAACGTGATGAGATA 5459
3600 lAspProGlnGlnValAlaGlnAlaLeuGlnGlnValAsnGlnAlaIysGlyAspLeuAs 3620
QY 5460 TCGAATAGTGAATAATAGTCAAGAAAGATGATGAAGATCTAAAAAATGTGTTCTGTACC 5519
3620 nGlyAsnHisAsnLeuGlnValAlaIysAspAsnAlaAsnThrAlaIleAspGlnLeuThr 3640
QY 5520 TTCAGATGATGATATCCATATGGGCAAAATCGCGATGAAGGTACTGCTATCAGTTTCT 5579

US-10-087-013-1 (1-10628) x US-09-919-172-29 (1-2125)

OY 3718 GGATGAAACCCAGAGGTACAGATTAACGAAAAAAGGAGATACG 3777
||| ||| :|||:||||
DB 19 GtlysalalysThrlleleleleuys-----ProArghsSer 32

OY 3778 GAATGTAACAGTAAATGATATA--CTTAAGAAAAAGATGAAGAAACAGTAGAA 3834
||| ||| :|||:||||
DB 33 AspCysProleuLysThrSerleProleuLysAlleCysAspTyrArglnleleu 52

OY 3835 -----GATTGCAATCCAAAAAGATAGATGATATCCCAT 3873
||| ||| :|||:||||
DB 53 llerhrlleTyrLysAspAspLysValleuAlaAspAsnSerHis--ArgAlaLys 71

OY 3874 TGGCAA-----TGGGAAATATTAATTTAGTGAAGACCCCGTGTGATG 3921
||| ||| :|||:||||
DB 72 TrpLysValleSerProThrleLysngluAlaMetVal-----ProSerValCysPhe 89

3922 -----CCCCPAGAAGACAAAAGTTATGCTACATTTCTGGCAAT-----GAT 3966
||| ||| :|||:||||
DB 90 ThrValProProAsnLysgluAlaVal-----AspLeuAlaAsnArgleleu 106

OY 3967 AATGAAATATAAATAAATTA-----CAATCAAGTATATTAAGAAGCT 4011
||| ||| :|||:||||
DB 107 GlnGlnTyrGlnAsnValleuThrleuTPrHisGlnSerHisleAsnMetLys----- 124

OY 4012 TTCATCAATCTGCAGCAGACAGAACATCTCTCATGTTATTTATTAAGTAAAGAT 4071
||| ||| :|||:||||
DB 125 -----SerValValSerTPrHisTyrleuLe----- 133

OY 4072 GGTGAAGAAATCAACTGCATAA-----GAATTAAGAAGGC 4110
||| ||| :|||:||||
DB 134 -----AsnGlnLleleAspArgleArgAlaSerAsnValleSerleLysThr-- 149

OY 4111 AAATTCCTCCCGCATTTTGTGATCCATGTTCTACACATTTGAGATTAAGATTTT 4170
||| ||| :|||:||||
DB 150 MetLeuProGlnLysleGlnleValleuSerAsnleuGlnSerArgPheGlnAspPhe 169

OY 4171 TTA-----TTTGAACAGATATTAACAAAGCTATGCT 4203
||| ||| :|||:||||
DB 170 LeuGlnAspSerGlnleuSerGlnAlaPheSerGlnSerAspIleThrGlnleuGlnLys 189

OY 4204 GAGGAACTAAACTAAAGACCAAAATAGATTCCTTTCAAAATGCTGACCAAAATCT 4263
||| ||| :|||:||||
DB 190 GlnValAsnValCysleGlnTyrGlnGlnleuLeuLysSerAlaGlnArgGlnGln 209

OY 4264 CCTATGAAAAACACGCCAANAATGCTGACAGAACATAGTCAT----- 4308
||| ||| :|||:||||
DB 210 GlnGlnGlnSerValTyrAsnleuTyrIleSerGlnValArgAsnIleArgleuArgleu 229

OY 4309 GAGATATGGAGCTATGCTATGCTACACTAGTAAATTTGGGCAAAAAAGATGATTTT 4368
||| ||| :|||:||||
DB 230 GlnAsnGlnAspArgleuIleArglnleuArgThrProleuGlnArgAspAspLeu 249

OY 4369 ACCGAAACTACGCTTACACACAGCTCAATTTAGTACAAAAACACACTTTGAGGAA 4428
||| ||| :|||:||||
DB 250 HisGlnSerVal----- 253

OY 4429 TTTGCCAAACGACCCAGTTTTCAGATGCTAACCGAATGCTAGACAGACTATTTGCTAT 4488
||| ||| :|||:||||
DB 254 -----PheArgIle 256

OY 4489 ACAGACAAAAATATTTGAAGATGTGACGAAAAATGTAACTCAATGACCAATTTGAAG 4548
||| ||| :|||:||||
DB 257 ThrGlnGlnGlnLysleuLysGlnleuGlnArgleuLys-----AspAspLeuGln 274

OY 4549 TGTGATACGATGTATATAGAAATGCGAGGACTACGTTAAATATATGAAAAAAGAAAA 4608
||| ||| :|||:||||
DB 275 ThrIleThr-----AsnLysCysGlnGlnPhePheSerGlnAlaAlaIleSerSer 291

OY 4609 GAGTGGATTCGA-----CAAGTAAATATTTACAGAGTGAACGCGACAAAAAA 4656
||| ||| :|||:||||

DB 292 Ser---ValProThrleuArgSerGlnleuAsnValleleuGlnAsnMetAsnGlnVal 310

OY 4657 AGATTCGATAGACAAACACATTTGCTATGCTTACAGACTATACT-----GG 4703
||| ||| :|||:||||
DB 310 LysIleSerMetSerSerThrTyrIleAspLysleuLysThrValleuValleuLysAs 330

OY 4704 AACGAATGCACAGATTTACTTGAACAGCAAAATTT--ACTGCTAGTGTGTGATTAAGCC 4760
||| ||| :|||:||||
DB 330 nThrGlnAlaAlaGlnAlaValleuValleuLysleuArgleuThrLysleuGlnGlnGln 350

OY 4761 TGGAAAGTCCCTGTGTGCTACAAAGAAATATACAAATTTGTA----- 4800
||| ||| :|||:||||
DB 350 aValleleleAsp-----LysAsnAsnleleGlnLysnleuIleSerThrleuLysGlnThr 368

OY 4801 -----GAAACAGGCTACATGATGCTGCGCAACCAATTTGGGCTG 4841
||| ||| :|||:||||
DB 366 PArgSerGlnValAspGlnLysArgleuValPheHisAlaLeuGlnAspGlnleuGlnLys 388

OY 4842 CACAAATTTATTTGAATGACGACAAATATACTAATTTGAGTAAAGAT----- 4893
||| ||| :|||:||||
DB 388 salAlaLysAlaIle-----SerAspLysMetPheLysThrTyrLysGlnArgAspLeuAspPhe 407

OY 4894 -----AATGCAAAAGATTAAGTAAAGGCGCAACACAGTGTCTATTAA 4937
||| ||| :|||:||||
DB 407 eAsPTrHisLysGlnLysAlaAspLysleuValleuGln-----Ar 420

OY 4938 GTGCAAAACAAAGGCTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4997
||| ||| :|||:||||
DB 420 GTrpGlnAsnValleHisValGlnleleAspAsnArgleuArgAspLeu-----GlnGlnIlele 439

OY 4998 TTTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5057
||| ||| :|||:||||
DB 439 LysSerleuLysTyrTyrArgAspThrTyrHisProleu-----AspAs 454

OY 5058 AGAAGTAAAGTGAAGAAATGGGTGCGAAGAAAGTGAAGAGAGTGGCGGACAGCAAGAG 5117
||| ||| :|||:||||
DB 454 PTrIleGlnGlnValleGlnThrThrleGlnArgLysIleGlnGlnAsnGlnProGlnAsnSe 474

OY 5118 GTCAATTTGGGTCAATTAATTAACAAAGAAAAA-----GAAAAAGA 5159
||| ||| :|||:||||
DB 474 LysThrleuAlaThrGlnleuAsnGlnleuLysMetleuValSerGlnleleuMetLys 494

OY 5160 GAAATTAACAAAGCTGGGATCCGACAAATATTTCTATGAGGTCGCCCTGTAGTCTAT 5219
||| ||| :|||:||||
DB 494 sGlnSerLysMetAspGlnLysGlnLysTyr-----AlaGln 506

OY 5220 GAAATTAAGTCTTT-----TATGATTTAAGCATATATATCTAGTATGATTA 5267
||| ||| :|||:||||
DB 506 uGlnTyrSerAlaThrValLysAspTyrGlnleuGlnThrMetThrTyrArgAla----- 524

OY 5268 TTTGGAAGATGAAAAACAAAGACGAGCAAAATTTGAAGAAATATTTAAACAAAAATG 5327
||| ||| :|||:||||
DB 525 -MetValAspSerGlnGlnLysSerProValLysArgArgMetGlnSerSerAlaAs 544

OY 5328 AACATCAGTTGGCAAGAGAGATGATCTACAGCAAAATCCCGGTAGTACTGCGGAA 5387
||| ||| :|||:||||
DB 544 PleuIleleleGlnGlnPheMetAspLeuArgThr----- 555

OY 5388 ATTTTCTGGAAGAAAAATGAAGATGTGTGGAACGCAATGATATGCGGGTACAAACG 5447
||| ||| :|||:||||
DB 556 -----ArgTyrThrAlaLeuValThrleuMetThrGlnTyrIleLys 569

OY 5448 TGTAGAGATGATGGAATATGGAATATGTCAGAA-----AGTGAAGAAATGATAA 5501
||| ||| :|||:||||
DB 569 s-----PheAlaGlnAspSerleuLysArgleuGlnGlnGlnleuLys 584

OY 5502 AAAATGTGTTGTGTAACCTTCAAGATGATATATCTATGGGAAAAATCGGAGTGAAG 5561
||| ||| :|||:||||
DB 584 sArgCys-----LysGlnThrSerGlnLys 592

OY 5562 TACTGCTGATCAATTTCTCGATGCTTTCGCAATGGGCTGAAGATTTTGCACAAATCA 5621
||| ||| :|||:||||
DB 592 sGlnLysArgleuSerAspLeu-----LeuGlnArgleuLys 603

```

OY 5622 AGAAAAAGGATTTGAGAAATTTGAGGGCGGTATGATTACTTGTGTGATTAATGA 5681
Db 603 salatThrValLeuGlu-----
OY 5682 AGATAAAAAGAAATGTACAGATGCGGTACACAAATATAAAAATTTATAGTAGTG 5741
Db 609 -----AsnSerLysLeuThrGlyLysIleSerGluLeuGluValMetValIleGluLe 626
OY 5742 GAAACCACTATGATAAAACAAATCAAAATATGTGAGAAATTAACCAAAATATATTC 5801
Db 626 uLysLysGlu-----LysSerTrpValGluI 635
OY 5802 CGACATCTCTGTGGCAAAAGATGAGAGAGCGCTCGGAATATTATAGCAAAATTAAT 5861
Db 635 uGluLeuProLysValArgGluIa-----AlaGluAsnGluLeuArgLysGluIlnr 653
OY 5862 AAAAAATTGTGAATAATTAAGTGGAGTTGTGAATTAAGTATGAAAGATGTGCCAC 5921
653 gAsnVal-----GluAspIleSerLe 660
5922 ACAGCGATTAATGATGATTAATGATCAAAATATATGCGCATCATTAAGAGATGAACCA 5981
Db 660 uGluLysIleArg-----Al 665
OY 5982 AGAAGTTGAAGAAATGTATTTGTCAATGCCAGACAGGTCCACACGTTAGCAAGGA 6041
665 agLysSerGluAlaLys-----GluTrpArgTrpArgI 675
OY 6042 AACACCGTACACAGCGGTATCACTGATATCAAAAGAGAGCGGATGCAAAAAAGACGAA 6101
675 uLeuGlu-----ThrIleValArgGluLysGluIaAlaGluArgLysLeu-- 690
OY 6102 AACAGCGCGCGCTACAAAAACAGCCGAAAAAGTGGAAATCTAACACAGAAATGCGAGC 6161
691 -----GluTrpValArgGluLeuThrIleGluAlaGluAl 702
OY 6162 ACAAAACAGACCCGACGAGCAGCA-----CAACAAC 6194
702 a-----LysArgAlaIleValIleGluLysLeuLeuAsnPheArgAsnGluLe 718
OY 6195 ACGAAAACAGACATCAACAGCAACACAGAAATCTGACGTGGGACAGATGGTAAAGC 6254
718 uGluLysAsnThrPheThrArgArgThrLeuGluAspHisLeu----- 732
OY 6255 CATCTTTGATTAACACAGATAGCAGGGGTGATAGAGGTTGTATCCAAAAACGTA 6314
732 ----- 732
6315 TGGACAATATCTTAATAGGGGTGTATTGTAGTAACTTAAGAAATGAATAATGCGAT 6374
732 ----- 732
OY 6375 ATGATGCTCTAGGAAAAAAATTTATGATAAATTAATTAACAATAT----- 6423
733 -----LysAspGlyAspLeuSerLeuAsnAspLeuGluGlnLysAsnLys 748
OY 6424 -TTAAATATGAACCTGAATAAGCGTGAACATGATTAATAAGCGCTTTATTAATG 6482
748 sleuMetGluLeuLeuArgArgLysArgAspAsnGlu--GluLysLeuLeuLysLeuI 767
OY 6483 TGCAGCAATAGAACTCAATTTTGTGGTTAAATATATATATTAATTAATCTCGACAGA 6542
767 eLysGluMetGluLysAspLeuAlaPheGlnLysGluValAlaGluLysGlnLeuLysG 787
OY 6543 AAATGAATTTGCAAAAATGGAACAATTCAGATGAATTTAAAAAGAAATATGATTAATACATA 6602
787 uLysGlnLysIleGluLeuGluAlaArgArgLysIleThrGluIleGlnLys----- 804
OY 6603 TGGTGAATTAAGATATGTTTTTGGAACTGATATTTCTAATGATTAATAAAATTAATTAAC 6662
805 -----ThrCysArgGluAsnAlaLeuProValCysPr 815
OY 6663 TGTAAAC-----AATAGTGTAAACAACATTTCTCAATGAATAATTAAGAA 6707
Db 815 oLleThrGlnAlaThrSerCysArgAlaValThrGlyLeuGlnGlnLysAspLysG 835
OY 6708 AAAACGAGAT-----AAAAAAGA 6728
Db 835 nLysAlaGluGluLeuLysGlnGlnValAspLeuLeuThrAlaAlaAsnArgLysAlaG 855
OY 6729 TGAAGAAATTAACATAAATATTTGGAGAAAAATTAATAATTTATTTGGAAAGCAATGAT 6788
Db 855 uGlnAspMetArgLysLeuThrTrpGluLeuAsn----- 866
OY 6789 ATATGATTAATCTTATCATCTCAGACAGCAAAAAAGCAAAAAAATTAAGATTAATTA 6848
867 -----AlaLeuGlnLeuGluLysThrSerSerGluLysAlaArgLeu----- 881
OY 6849 CCAGTACAAATGACATGACAAACCTGACCGCTTCCTGGAACAGTTTGTAAAAAGCCCCCA 6908
882 -----LeuLysAspLysLeuAspLeuThrAsnAsnThrLeuArg 894
OY 6909 ATTTTGAAGATGTTTACAGAAATGGCAGAGAAATTTGTAAATGAAGCAAGACAGTT 6968
894 gCysLeuLysLeuGluLeuGlu-----ArgLysAspGlnAl 906
OY 6969 GTTAAAA-----TTGAGCGCGCGCTGTAAAGAAATATAGTGTAAATGG 7010
906 agLysLysGlyTrpSerGlnGlnLeuArgGluLeuGlyArgGlnLeuAsnGlnThrThrG 926
OY 7011 TAGTAATGACGTGAAGACACAGAAATGTGCAGAGCGCGTGTAAACATATCAAAATTTAT 7070
926 YLysAlaGluGluAlaMetGlnGluAlaSerAsp-----Le 938
OY 7071 TAAAGATGGAAGAACTGAATTTGAAGACAAAGCAAAAGTCTAAAAAGATTAAGATGG 7130
938 uLysLysIleLysArgAsnTrpGlnLeuGlnLeuGlnSerLeuAsnHisGluLys--G 957
OY 7131 CAAAAAGTAAAGATATATCTTACTGAAGAGACATGAGAGCAACATAGTGTCTCA 7190
957 YLys-----LeuGlnArgGluValAspArgIleThrArgAlaHis 970
OY 7191 T-----GAAATTTTAAACATGAATAATTAAGAAATTAATGTTGCGCAA 7229
970 sAlaValAlaGluLysAsnIleGlnHisLeuAsnSerGlnIleHisSerPheArgAsp 990
OY 7230 TAAAGATTTGTTCT-----TGTATGCAAAAACCT-----TCTTCAACAAT 7268
990 uLysGluLeuGluArgLeuGlnIleCysGlnArgLysSerAspHisLeuLysGluIlnr 1010
OY 7269 ACCAAAAACACACAAACA-----TCACAATCATCCGATGTAATGATAT 7313
1010 eGluLysSerHisGluGlnIleLeuGlnHisAsnIleLysAlaGluLysGluAsnAspLys 1030
OY 7314 GCCAGATCGCTGATTAATGTTCTGTAAGAATTTTAACAAGTGT--GAGTGTCTGAAT 7370
1030 sIleGlnAlaGluLeuAsn-----GluLysLeuGluLysSerAsnGluCysAlaGluMe 1047
OY 7371 TTCAAAAAGGAGTCTATGATTCATCAAAAATAATTAAGTGAACCTTAATGATGA 7430
1047 tLeuLysGlnLysValGluGluLeuThrArgGlnAsnAsnGlnIlnrLysLeuMetCysG 1067
OY 7431 TTCTGTAGAGAAACGCA----- 7449
1067 nArgIleGlnAlaGluSerGluAsnIleValLeuGlnLysGlnThrIleGlnArgC 1087
OY 7450 -----TATTA 7454
1087 sGluAlaLeuLysIleGlnAlaAspGlyPheLysAspLeuLeuArgSerThrAsnGluH 1107
OY 7455 TTTATCTAAAGACAGAAATAATATATGATATCTTGAAGCAAAATTAATATACCT-- 7512
1107 sleuHisLysGlnThrLysThrGlnGluAspPheGlnArgLysIleLysCysLeuGluG 1127
OY 7513 -----ATTGAGTCTACAAAGAAAG--GAAAGTAAAAA 7544

```

Db 1127 uaspleuAlaLysSerGlnAsnLeuValSerGluPheLysGlnLysCysAspGlnGlnAs 1147
 QY 7545 TAGTTGGCAATAATTAATCTTGGATCCCTTAAGAAACCTTATGACCTGATTAATATAT 7604
 Db 1147 nilellelleGlnAsnThrLysLysGlu-ValArgAsnLeuAsnAlaGluLeuAsnAlas 1167
 QY 7605 AGGAGAGAGAAACCTTGTGAAAATAGAGAGAGAAATCGTTTAAAGTGATGATT- 7657
 Db 1167 erLysGlnLysArg---ArgGlyGlnGlnLysValGlnLeuGlnGlnAlaGlnValG 1186
 QY 7658 ----ATGAATGGAATGTTTACAAAATTCAAAGTCTTATGACGAGAGAAA- 7702
 Db 1186 lngluleuAsnAsnArgLeuLysValGlnAspGluLeuHisLeuLysThrIleGlnG 1206
 QY 7703 ----AAAGATATGTCTACCTCCAGAGAGAAC- 7736
 Db 1206 lngluleuThrHisArgLysMetValLeuPheGlnGlnLysSerGlyLysPheLysGln 1226
 QY 7737 GTGCTTAAGCAATTTAGATGAATTAATTAAGCACTTAAGGATAGTAATATCTCT 7796
 Db 1226 erLla-GluLysPheArg-Lys---LysMetGluLysLeuMetGluSerLysValIleTh 1244
 QY 7797 AAAAATGTTCTGCAATGCAAGCAATGAGAAATGACATTAATTAATAAACTTCAACT 7856
 Db 1244 rgluAsnAspLleSerGlyLeuArg- 7857
 QY 7857 AGAGACGGGTGCCCAATGAAAT-CCAAATGTGATCTATGAAATATAGTTTCCG 7910
 Db 1260 ngluAsnSerArgLacGlnGlnLysAlaLysLeuGlySerGluThr- 1274
 QY 7911 TGATCTGGGTGACATAGTTAGAGCAACATATGTTACGAATGTGGTACTTACCTCC 7970
 Db 1275 -AsnIleLysGlnLeuGlnArg- 1281
 QY 7971 CGTAGAAATTAATTAATATAGTTTGAATACATATATGAAATGAGAAATAAATA 8030
 Db 1281 - 1281
 QY 8031 TAAAGGTAGAAATTAATTAACAGCATGTAACAACGTTTCGTTCTGCTGGGATCTTA 8090
 Db 1282 -GlnLeuGlnGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1294
 QY 8091 TAGAAAAGATATTGGAAGCAATGACGTGCAAGCAACAGAGATGCAAACTTTTAC 8150
 Db 1294 nHisMetGlnAlaAsnHisLysLeuLysCysGlnLysLeuGlnAspGluLeuLeuIle 1314
 QY 8151 AAAAGAGAAATGATGATTTGACGCAATACATTAATACAGATAAAGTGTGACATTA 8210
 Db 1314 nLysArgGlnValGlnAsnLeuLysGln- 1324
 QY 8211 GCAGCATGCACTGTGATGATTAATACCTCAACGTTTCGATGATGACATGATGTC 8270
 Db 1324 sMeLAspGlnGlnIleLysGlnHis- 1332
 QY 8271 TGAATTTATTGTAAGCACTGATGAGAAATGGAATAATTAATAATCATGTGATCA 8330
 Db 1333 -GlnHisGlnLeuValLeuLeuGlnCysGlnIleGlnLys- 1351
 QY 8331 CTGTAAACATCTGCAGATGCAAGATGATTAATGTAATAATTAAGTGCACAGTGTA 8390
 Db 1351 pCys- 1362
 QY 8391 AACGAGATGTCAGAAATAT- 8409
 Db 1362 LysGlnCysGlnHisSerGlyLysLeuSerSerArgAsnThrGlyHisLeuHisProth 1382
 QY 8410 ----AAAAATTTGTTTAAATGGAATCTGATTCGATATCAATCAATTAATACAA 8465
 Db 1382 rProArgSerProLeuLeuArgThrProGln- 1398
 QY 8466 AGAATGTATGAACAACCAATATATACAAAATCTGACTTATGATCATCTCAAAATTT 8525

Db 1398 uGlnLysThrGlnHisArgValValGlnGlnIlePro-----LysGlnValGln---Ph 1415
 QY 8526 TGTACAAAAGTTGAAAACCTTTTAAAGTGAATGTTCTGTTGAGAGCTTTTCAATATCT 8585
 Db 1415 eGlnProProGlyAlaProLeuGlnLysGlnLysSerGlnGlnCysThrSerGlnLysThr 1435
 QY 8586 TCATGAACAGATAGTCTGTTGAATTAATTAATTAATGAATAATGATGCTTCTCC- 8640
 Db 1435 eSerGlnThrSerThrGlnLeuGlnIleThrPheAspGlnThrAsnProIleThrArgLe 1455
 QY 8641 ----AATATGCAACATATGCTTTCGAGAGAAACA-CCAAAAGTAA 8681
 Db 1455 uSerGlnIleGlnLysIleArgAspGlnIleAlaLeuAsnSerArgProProValArgL 1475
 QY 8682 TAAAGAA- 8706
 Db 1475 rgluAspAsnAlaCysGlnMetGluLeuValLysValLeuThrProLeuGlnIleAlaL 1495
 QY 8707 - 8723
 Db 1495 sAsnLysGlnLysAspMetHisThrGlnValThrThrLeuLysGlnGlnLysAsnProVa 1515
 QY 8724 GATTAATTTGCTACCGATCAAAACAAAGATGATGAAGAA- 8774
 Db 1515 lProSerAla- 1534
 QY 8775 TTTTACCTCTGCTCGAGCAATGATTAATTAATCTGTAATTTGCAACCGATACCT 8834
 Db 1534 sGlyAspPheLeuLysGlnLysLeuGlnProGlnThrPheGlnAsnProLysAspH 1554
 QY 8835 TGTTCTTAATAGTTCAGATTAAC- 8859
 Db 1554 sAlaCysSerValArgAspArgLysPheGlnGlnLysLeuArgHisThrValThrAl 1574
 QY 8860 - 8883
 Db 1574 aArgGlnLeuValGlnAlaLysLeuLeuAspMetArgThrIleGlnGlnLeuArgLeuG 1594
 QY 8884 - 8909
 Db 1594 yLeuLysThrValGlnGlnValGlnLysThrLeuAsnLysPheLeuThrLysAlaThrSe 1614
 QY 8910 CACTGCATTAATTAATGAAGAGGTGATTAAGAA- 8943
 Db 1614 rIleAlaGlyLeuLysLeuGlnSerThrLysGlnLysLleSerPheAlaSerAlaIle 1634
 QY 8944 - 8988
 Db 1634 uArgIlellellelleAspLysMetValAlaLeuAlaPheLeuGlnAlaGlnAlaIle 1654
 QY 8989 - 9024
 Db 1654 yPheIlellelleAspProIleSerGlnThrLysSerValGlnAspAlaValLeuLysG 1674
 QY 9025 - 9044
 Db 1674 yValValAspProGlnPheArgIleArgLeuLeuGlnAlaGlnLysAlaAlaValGly 1694
 QY 9045 TAGTTATGAGATTAATCCGATTAATTAAGAGAACTGATATGACACTCATTAAT 9104
 Db 1694 rSerLys- 1697
 QY 9105 TGAATAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9164
 Db 1697 rSerLysThrLeuSerValPhe- 1712
 QY 9165 GTGGGAAAATTAATAGACGTAGATATGCGACGCTATGTTGTGGATATTAATTTGCTAC 9224
 Db 1713 - 9225
 QY 9225 TTTCAAAAGTAAACATTAATGATGAAGATGCTGTCATTAATCAAGAGATGAAGAACTAATCA 9284
 Db 1727 rGlnGlyValIleAspProValArgGlyIleArgValProProGlnIleAlaLeuGlnG 1747

```

QY 9285 GTTCTGTTGGTTAATTGAATGGCAAGCAGTGTAGAAAGAAAGATGAG 9344
Db 1747 nglyleu-----leuansnmlalleuenglneleuhsiglupr 1761
QY 9345 TGATTCATTAACAAATGCTCGTTCAAGCAGATATTTTGAAGC----- 9396
Db 1761 oserseerantprargvalphepro-----AsnProasnsnlysglnalleuTytr 1779
QY 9397 -TCAGATTTATTAAGCAACCTGATGTCAGATTTTGAATTTGAATATATTACTGAA 9455
Db 1779 rsergluleuLeuargmet-----CysValPheasprvalgluserglncyspheleuph 1797
QY 9456 TATATTGTAATAAATTAACATGGAATCTAAATATAAATATAAGCAATTAAGATCA 9515
Db 1797 eProPhegluargvalgaasnIleSerAsnleuasnVallys-----LysThrHl 1813
QY 9516 ATCTTACGATTAATAGCAATTAACCA---TCTGAAGAAATTTGACATATATAA 9572
Db 1813 sargIleSerValValasprThrlysthrlysergluleuThrValTyrglu----- 1830
QY 9573 ATCAAGATTTCTCAATGCCGCTTGGAGTTAAATGATATAATGAATAGTTACAGAAC 9632
Db 1831 -----AlaphegnIArgasnleuIleglulysserIleTyrglu 1844
QY 9633 AAAAAATTAATGAATATATGAAATTCAGAAAGTACTAAAAATTAATATCTGCTTATA 9692
Db 1844 uleuSerclnglInItyrglnItyrlysgluIalmet----- 1856
QY 9693 TTTTGTGAAGANGAACAACACAAATCATGTACTAGATGGA----- 9735
Db 1857 -PhePhegluserTyrclyHisserHisMetleuThrAspThrlysthrclyleuHl 1876
QY 9736 -----AATATAAAGAAAGAA---GAGCAACAGATTCCTCAAGACCTGATTTCTTAC 9788
Db 1876 sPheasnsIleasngluIalIleeglInglTythrIleasprlysalaleuVal----- 1893
QY 9789 ACCCATGATGATCTTCTCATACACACCTTATTCACACATCGATGACA----- 9843
Db 1894 -----LyslysthrlysglIngluleuIleThrleuThrcluleuIleaspsse 1909
QY 9844 -----CAATATGATTCCTTAATAATGATATATATGAAAGTATCTGTTGTTAT 9893
Db 1909 rleuLeuSerargLeuValprolystlysasp---LeuHisSerProvalalglTytyr 1928
QY 9894 TGTATCGGCGTATAGT-----TTGATACGCGTTCATTCAGAAAGAAATTCATATC 9947
Db 1928 pleuThrIleaserglygluIargIleSerValleuIySalaseraArgasleuValas 1948
QY 9948 GTCGTGGACTTGTGCGCATACTGATATCCCGCAAGAGATGGAATGCCCTACGTT 10007
Db 1948 pArgIleThrIalaleuArgCysleuclualaglnValSerThrlyglyIle----- 1965
QY 10008 GGAATCCAAAAATAGTACATACCATATAGAGTGTCCATATAAGCAAAACATATAT 10067
Db 1966 -----IleasprleuThrlystlysTytyr 1975
QY 10068 AATATGAGAGAGATACTAGTGAAGATGAATATAATATGCGGACTATATCTCTC 10127
Db 1975 gValIleaglualaleuHlsarglyLeuValasprgluleuPhe----- 1999
QY 10128 TGATATTACTTCATCCGAAAGTGAATGAGAAATTTGATATATATATATATATAT 10187
Db 1990 -----AlaeglInleuValargInCysgluleuValIleThrlyglIleghIstpr 2006
QY 10188 AGGTAGTCTTAATTAACATGATGATGAGATGATA 10224
Db 2006 oIleThrAsnlysmetSerValValgluIalVal 2018

```

RESULT 15
US-10-087-464-10

; Sequence 10, Application US/10087464
; Publication No. US2003059436A1

```

; GENERAL INFORMATION:
; APPLICANT: Chishli, Athar
; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuerong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses There
; FILE REFERENCE: S1237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-087-464-10

```

Alignment Scores:

```

Pred. No.: 2,01e-10 Length: 1639
Score: 289.00 Matches: 330
Percent Similarity: 33.65% Conservative: 240
Best Local Similarity: 19.48% Mismatches: 605
Query Match: 1.49% Indels: 520
Db: 9 Gaps: 76

```

US-10-087-013-1 (1-10628) x US-10-087-464-10 (1-1639)

```

QY 4474 GACGACTATTCCTAT-----ACAGCAAAAAATTTGAAAGATGTG 4515
Db 198 AsnAsprlyrcysglnIleProPheasnsleuIleArgIleasngluLeu---AspVal 216
QY 4516 CAGCAAAATGT-----AAGTCAATGACCAATTAAGATGATGATACGAA 4560
Db 217 leuIySalaleuValPheglTyArglyspProleuaspnsnIleuysAspVal--- 235
QY 4561 TGTATAAAGAAATGCGAGGATGATTAATATATGAAAAA----- 4602
Db 236 -----GlyysMetgluAsprlyrIleIySalasnslysthrIlegluaenIleasn 253
QY 4603 -----AAAAAGATGATTCACAGATTAATATATTAACAGATGAACG 4647
Db 254 gluleuIleaglIngluserIystrlysthrIleasprlyasnsnlysalahrllysglu 273
QY 4648 GACAAAAAAGATTCGATAGCAACACATGTGTATAGTTACAGACTATACCTGGAACG 4707
Db 274 glulyslystlyseuTyrglnIalaglnTyraspleuSerIle----- 287
QY 4708 AATGCAACAGATTACTGAAACAGAAATTTACTGTAAGTTGCGATGAAGCTGGAAGT 4767
Db 288 -----TyrasnlysglnleuclInluIahlsnsleu 298
QY 4768 GCGCTGTGTGTAAGAAATATATACATTTGAAAAAGAGCTTATGATGCGGAC 4827
Db 299 IleSerValleuclInluysargIleasprThrleuIylys----- 311
QY 4828 AAACATTTGGGTGACCAAAATTTATTGAAAATGACGCAAAATATACATTTTCAGT 4887
Db 311 ----- 311
QY 4888 AAGATTAAGTGAAGATTTAGTAAGGAGCAACACAGGCTTATTAAGTGGCAAAAC 4947
Db 312 AsngluAsnIleIysgluleuLeuaspIyIleasn-----gluIleIySalasn 327
QY 4948 AAGGTCTATATATACATAATGAAAGATTTGAAAGATGACTGAAGATGCTCTTCTCT 5007
Db 328 ProProIalIasnsrgIyasn-----ThrProasnthleu--- 340
QY 5008 CGTGCATGATGATATGTTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 5067
Db 341 -----LeuaspIySalasn-----Lys 345

```


OY 5068 GATGAAAATGGGTTGGCAAAAAGATTGATGAGTGGCGCAACGGAGGTACATATTG 5127
 DB 346 LysIleGluGlnHisGluLysGluIleLysGluIleAlaLysThrIleLysPheAsnIle 365
 OY 5128 GGT-----CAATCTACAAAAGAAAAAGAAAGAGAG 5160
 DB 366 AspSerLeuPheThrAspProLeuGluLeuGluTyrTyrLeuArgLysAsnLysAsn 385
 OY 5161 -----AAATTAATAAGCGCATGGCAAAATATCTATGAGTGGCGG 5205
 DB 386 IleAspIleSerAlaLysValGluThrLysGluSerThrGluProAsn---GluTyrPro 404
 OY 5206 CCTGTAGTGTATGATAATATAGTTTATGATTAAGAGATATATCTAGTATGAT 5265
 DB 405 AsnGluValThrTyrProLeuSerTyrAsnAspIleAsnAlaLeuAsnGluLeuAsn 424
 OY 5266 AATTGGGAAGAT-----GAAAAACAAGACCGAGAAATTTGAAAGAAATATT 5316
 DB 425 SerPheGlyAspLeuIleAsnProPheAspTyrThrLysGluProSerLysAsnIleTyr 444
 OY 5317 AACCAAAATGGAACATCACTGTGCAAGAGATGATAGTACAGAAATCCCGTAGT 5376
 DB 445 ThrAspAsn----- 447
 OY 5377 ACTGGCGCAAAATTTTCTGAGACGAAATAAGATGTGTGTGCAACGCAATGATATGC 5436
 DB 448 ---GluArgLysLysPheIleAsnGluIleLysGluLysIle-----LysIleGlu 463
 OY 5437 GGGTCAACAAACGTGTAGATGATGAAATAGTGAATAGTGCAGAAAGTGTGAAGT 5496
 DB 464 LysLysLysIleGluSerSerLysLysSerTyrGluAspArgSerLysSerLeuAsnAsp 483
 OY 5497 CTAAAAAATGCTGTCTGACCTTCAGAGATGATTAATCTATGCGGAAAAATCCGCAT 5556
 DB 484 IleThrLysGluTyrGluLysLeuLeuAsnGluIleTyrAspSerLysPheAsnAsnAsn 503
 OY 5557 -----GAAGTACTGCTATCAGTTCTT----- 5580
 DB 504 IleAspLeuThrAsnPheGluLysMetMetGlyLysArgTyrSerTyrLysValGluLys 523
 OY 5581 -----CGATGCTTCCGCAATGGCGTGAAGATTTTGCAGAAATTAAGAAAG 5628
 DB 524 LeuThrHisHisAsnThrPheIleAsnTyr-----GluAsnSerLysHisAsnLeuGlu 541
 OY 5629 GAATTCGAGAAATGCTAGCGCGCTGATGATTAATCTGTCGTATGATGAGATATA 5688
 DB 542 LysIleThrLysAlaLeuLysTyrMetGluAspTyrSerLeuArgAsnIleValGlu 561
 OY 5689 ACAGAAATGTACAGATCGGTACACAAATTAATAAATTTATTAAGAGTGAAGACA 5748
 DB 562 LysGlu-----LeuLysTyrTyrLysAsnLeuIleSerLysIleGluAsn 576
 OY 5749 CAGTATGAAAAACAATCAAAAATATGCTGAGAAATTAAGACAAAATATATCCGAGCAT 5808
 DB 577 GlnIleGluThrLeuValGluLysAsnIleLysLysAspGluGluLeuPhe---GluLys 595
 OY 5809 CCTGTGCAAAAGATGCAGAGACGCTCGCAATATTAGACAAACAATTAATAAATTT 5868
 DB 596 LysIleThrLysAsp----- 600
 OY 5869 TGTGAATATAAGTGTATGAAAGATGTGCCACA 5922
 DB 601 ---GluAsnLysProAspLysIleLeuGluValSerAspIleValLysValGluVal 619
 OY 5923 CAGCGATTAACGTAGTATAGTCAAAATATGCCCCGATCATATAGACATGAACAAA 5982
 DB 620 GlnLysValLeuLeuMetAsnLysIleAspGluLeuLysLysThrGlnLeuIleLysLys 639
 OY 5983 GAAGTTGAAGAAATGTATATGTCAGATGCCAGAGTCCACAGCTGTAGCAAGGAA 6042
 DB 640 AsnValGluLeuLysHisAsnIleHisValProAsnSerTyrLysGluGluLysAsnLysGln 659

OY 6043 ACACGCTCACACGGGTATCACTGATATCAAAAAGCAGCGCATCGAAAAAGAA----- 6096
 DB 660 GluProTyrTyrLeuIleValLeu-----LysLysGluIleAsp 672
 OY 6097 ---GCGAAAACAGCGCCCTACAAAACAGCCGAAAAAGTGGAAAAATGTACACAGAA 6153
 DB 673 LysLeuLysValPheMetPro-----LysValGluSerLeuIleAsnGlu 687
 OY 6154 ATCGGACGACAAACAGAAACCGAGAGACGACACAAACAAACGAAACGAAACATCAACA 6213
 DB 688 GluLysLysAsnIleLysThrGluGluGlnSerAspAsnSerGluProSerThrGluGly 707
 OY 6214 GCACACACACAGAAATCTAGCTGGCACATGTAAAGCCATCTTCTCGAATAAACCA 6273
 DB 708 GluIleThrGlyGlnIle-----ThrThrLysPro 717
 OY 6274 GATAGCAGGGGTGCA-----ATAGAGGTTGTAATCCAAAAACGTATGACAAATTCCT 6327
 DB 718 GlyGlnGlnAlaGlySerAlaLeuGluGluLysAspSerValGlnAlaGln 735
 OY 6328 AATGCGGTTGATGTAGTAAAGTAAAGAAATGAATAAGCATATGCTATGCCT----- 6384
 DB 736 -----GluGlnLysGlnAlaGlnProProValProVal 748
 OY 6385 CCTAGAGAAAAAATATATATATTAATATATATATTAATTAATTAATGAACTGAAAT 6444
 DB 749 ProGluAlaLysAlaGlnValProThrProProAlaProValAsnAsnLysThrGluAsn 768
 OY 6445 AAGCGTGAACATGATTAATAAGAGCTTTTATTAATGTGACAGCAATACAACTCAATT 6504
 DB 769 ValSerLysLeuAspTyrLeuGluLysLeuTyrGlu-----PheLeuAsnThrSerTyr 786
 OY 6505 TTGTGGTTAAATATATTAATTAATGAAATCCTGCAGCAAAATGAA-----TTGCAAAAT 6558
 DB 787 IleCysHisLysTyrIleLeuValSerHisSerThrMetAsnGluLysIleLeuLysGln 806
 OY 6559 GGAACATTCACAGATGAATTAATAAGAAATATGATTAATACATATGATGATTAAGAT 6618
 DB 807 TyrLysIleThrLysGluGluGluSerLysLeu-----SerSerCysAspProLeuAsp 824
 OY 6619 ATGTTTGTGGAACATGATTTTCTATGATTAATAAATTAATA-----ACT 6663
 DB 825 LeuLeuPhe-----AsnIleGlnAsnAsnIleProValMetLysSerMetPheAspSer 842
 OY 6664 GTAACAAATAGTGTACACACCAATTCGAATGAATAAATAAGAAAAA----- 6711
 DB 843 LeuAsnAsnSerLeuSerGlnLeuPheMetGluIleTyrGluLysGluMetValCysAsn 862
 OY 6712 -----CAGCATAAAAAAGAT-----GAGAAATTAAGTAAATA 6747
 DB 863 LeuTyrLysLeuLysAspAsnAspLysIleLysAsnLeuLeuGluGluLysVal 882
 OY 6748 TTTTGGAGAAAAAATTAATTAATTTATTTGGAGAAAGATATATATGATTAATCTATCAT 6807
 DB 883 -----SerThrSerValLysThrLeuSerSerSerMetLeuProLeuSer----- 898
 OY 6808 CTCACAGACGAAACGAAAGAAAAAATTAAGATTAATTAACAGTACATGATGATGACC 6867
 DB 899 LeuThrProGln---AspLysProGluValSerAlaAsnAspAspThrSerHisSerThr 917
 OY 6868 AAACGTAGCGCTTCCCTTGAGAGATT----- 6894
 DB 918 AsnLeuAsnAsnSerLeuLysLeuPheGluAsnIleLeuSerLeuGluLysAsnLysAsn 937
 OY 6895 -----GTAAAAAGCCCAATTTTGTGAGATGCTTACAGAA 6930
 DB 938 IleTyrGlnGluLeuIleGlyGlnLysSerSerGluAsnPheTyrGluLysIleLeuLys 957
 OY 6931 TGGGCGAGAGAAATTTGATTAAGAGAGAGAAACAGTCTTAAAAATGAGCGGCGCTGT 6990
 DB 958 AspSerAspThrPheTyrHisGlnGluSerPheThrAsnPheValLysSerLysAlaAsp----- 976
 OY 6991 AAGCAATATGAGTGTATGATGATTAATGACGCTAAGACACAGAAATGTGACAGCGCTGT 7050

```

Db 977 -----AspIleasnSerLeuasnAspGluSerLysArgLysLysLeuGluLysAsp 993
QY 7051 GTA-----ACATATCA-----AATTTTATTAAGAGTGGAAACT 7086
Db 994 ILeasnLysLeuLysLysThrLeuGlnLeuSerPheSerPheLysLysLysLysLeu 1013
QY 7087 GAATATGAAGACAAAGAGAAAGTTCAAAAGATTAAGATGCAAAAGTAAAGAT 7146
Db 1014 LysLeuGlnLys-----LeuPheAspLysLysLysThrValGlyLysLys 1029
QY 7147 TATCCTCTCTGTAAGAGACATAGAGCAACATGCTCTATGAATTTTAAACATG 7206
Db 1030 -----MetGlnIleLysLysLeuThrLeuLeuLysGluGlnLeuGlnSer 1044
QY 7207 AAATTA-----AAAGAATATGTGTCGAATAGGATGTTCTGTATG 7248
Db 1045 LysLeuasnSerLeuasnAsnProLysHisValLeuGlnAsnPheSerValPhePheasn 1064
QY 7249 CAAAAACCTCTCACAACCTACCAAAACACAAACAAATCAATCATCCGATGCT 7305
Db 1065 LysLysLysGluAlaGlnIleAlaGlnIleThrGluAsnThrLeuGlnAsnThrLysIleLeu 1084
QY 7306 -----AATGATATGCGAATCGCGATTTAT 7332
Db 1085 LeuLysHisTyrLysGlyLeuValLysTyrTyrAsnGlyLysSerProLeuLysThr 1104
QY 7333 GTTCCGGAAGATTTTACAAGTGTGCTCTGAACCTTCAAAAAGGATCTATGAT 7392
Db 1105 LeuSerGlnGlnSerIleGlnThrGluAsnTyrAlaSerLeuGlnAsnPheLysVal 1124
QY 7393 CATACAAAAAATTTACTAACCTAAATACCTAGATATGTGTAGAGAAACACACAT 7452
Db 1125 LeuSerLysLeuGlnGlyLysLysLysAsnLeuAsnLeuGlnLysLysLysLeuSer 1144
QY 7453 TATTTATCTAAAGACACAAATAATATGATATTTACCTGGAAGAAAAATTTATCCT 7512
Db 1145 TyrLeuSerSerGlyLeuHisHis-----Leu 1153
QY 7513 ATTGAGTCTACAAAGAAAGAAAGTAATAATAGTGTGCTAATTAATCTTGCAT 7572
Db 1154 ILeAlaGlnLeuLysGlnValIleLysAsnLysAsnTyrThrGlyAsnSerPro 1171
QY 7573 CCTAAGAAACCTTATGACACCTGATTAATATATAGAGAAACCCCTGTGAAATAGA 7632
Db 1171 ----- 1171
QY 7633 GAAGAAATCGTTTACGTACATTAATGAATGGAATGTTACAAAATTCAT 7692
Db 1172 SerGluAsnAsnThrAspValAlaAsnAlaLeuGlnSerTyrLys-----LysPheLeu 1189
QY 7693 CAGGAGAAAAAGAGTGTGTACCTCCAGAGAGAACATATGTGCTTAAGAAATTA 7752
Db 1190 ProGlnGlyThrAspValAlaThrVal-----Val 1199
QY 7753 GATGAATTAATAATTGAAGACTTAAGATATTAATCTCTCAAAAATGTTCTGCGA 7812
Db 1200 SerGlnSerGlySerAspThrLeuGlnGlnSerLys-----ProLysLysProAlaSer 1217
QY 7813 ACT-----GCACGAAATGAAGAAATGACATTAATTAATAAACTCAACTCGAG 7860
Db 1218 ThrHisValGlyAlaGlnSerAsnThrIleThrThrSerGlnAsnValAspAspGlnVal 1237
QY 7861 AACGGGTGCGCAATGCAATATGTGATCTATGAATTAATTAATGTTGCGGATCGGT 7920
Db 1238 AspAspValIleIleValProIlePheGlyLysSerGlnGlnLysAspTyrAspAspLeuGly 1257
QY 7921 GACATAGTTAGAGAACAGATATGTACGAATTTGTTGTTACTCCCGTGAAGATA 7980
Db 1258 GlnAlaValThrGlyGlnAlaVal-----ThrProSerValIleAspAsnIleLeuSer 1275
QY 7981 AAATTAATTAAGCTTTTGAATACATATAT-----GGAATAATGAGAAAT 8025
Db 1276 LysIleGlnAsnGlnTyrGlnValLeuTyrLeuLysProLeuAlaGlyValTyrArgSer 1295
QY 8026 AAAAATTAAGTAGAAATTAATTAACAAGATGTACAAACGTTTGTGCTGGTGGAT 8085
Db 1296 LeuLysLys-----GlnLeuGlnAsnAsnValMetThrPhe-----Asn 1308
QY 8086 GCTATATAGAAAAGATATTTTGAACCAATGACGTGCACCAAGACAGATGCAAAACT 8145
Db 1309 ValAsnValLysAspIleLeuAsnSer----- 1317
QY 8146 TTAGAAAAAGAAATGATGATGATTGACGATTAATTAATTAAGATNACTGTGCA 8205
Db 1318 ---ArgPheAsnLysArgGlnAsnPheLysAsnVal---LeuGlnSerAspLeuIlePro 1335
QY 8206 CATTAAGACGATCCACCTGTGATGATTAATATACCTCAACGGTTCGATGATGATGAA 8265
Db 1336 TyrLysAspLeuThrSerSerAsnTyrValLysAspProTyrLysPheLeuAsnLys 1395
QY 8266 TCGTCTGAATATTAATTAAGCACTGATGAGAAATTTGCAAAATTTAAAAATCATGT 8325
Db 1356 -----GlnLysArgAspLysPheLeuSerTyr 1365
QY 8326 GATCAGTGTAAA-----ACATCTGACAGATGCAAGATGAT-----TAT 8364
Db 1366 AsnTyrIleLysAspSerIleAspThrAspIleAsnPheAlaAsnAspValLeuGlyTyr 1385
QY 8365 GATGAATTAATGATGAAACAGTGTAAACGAGATGTCACAGATTAATAA----- 8413
Db 1386 -TyrLysIleLeuSerGlnLysTyrLysSerAspLeuAspSerIleLysLysTyrIleAs 1405
QY 8414 -----AATTTCTTCTTAATGAAGAAATCTATTCGAT 8445
Db 1405 nasPlyGlnGlnGlnLysAsnLysLysLysLysLysLysLysLysLysLysLysLys 1424
QY 8446 ATACATCAAAATTAATCAAAAGAA----- 8469
Db 1424 TyrLysThrValAsnAspLysIleAspLeuPheValIleHisLeuGlnAlaLysValLe 1444
QY 8470 -----TTGATGACAAACCA-----ATATATACAAAATCTCTACTATGATCATGTTCA 8519
Db 1444 uAsnTyrThrTyrGlnLysSerAsnValGlyValLysIleLysGlnLeuAsnTyrLeuLys 1464
QY 8520 AAATTTGTACAAAGCTGAAACCTTTTAAAGTGAATGTTCT-----GTTGA 8567
Db 1464 sthIleGlnAspLysLeuAlaAspPheLysLysAsnAsnAsnPheValGlyIleAlaLys 1484
QY 8568 GAGCTTTCTGAATATCTCAT----- 8589
Db 1484 PheSerThrAspTyrAsnHisAsnAsnLeuLeuThrLysPheLeuSerThrGlyMetVal 1504
QY 8589 ----- 8589
Db 1504 PheGluAsnLeuAlaLysThrValLeuSerAsnLeuLeuAspGlyAsnLeuGlnGlyMet 1524
QY 8589 ----- 8589
Db 1524 tLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGlnAsnSerGlyCysPhe 1544
QY 8590 -----GAAACAAATTAAGT-----TTGAATTAATAA----- 8616
Db 1544 eArgHisLeuAspGluArgGlnGlnLysLysCysLeuLeuAsnTyrLysGlnGlnGlyAs 1564
QY 8617 -----TTTAATGAATAATGATGTTCTTCCAAATATACG 8648
Db 1564 PLYSCysValGluAsnProAsnProThrCysAsnGlnAsnAsnGlyCysAspAlaAs 1584
QY 8649 AACATATGCTTTCGAAGAAACCAAAAAGT-----TATAAGAACGCTTGCAGTTGTAC 8702
Db 1584 palLysCysThrGlnGlnAspSerGlySerAsnGlyLysLysIleThrCysGlnCysTh 1604
QY 8703 ACTACCTTGAAGAAATTCATGATTAATTTGCTACCGATCAAAACAAAGATGATGA 8762
Db 1604 rLysProAspSerTyrProLeu-----PheAspGly----- 1614

```

Mon Apr 28 13:47:29 2003

us-10-087-013-1.n2p.rapp

Page 41

```
OY  8763  GGAATTACAACCTTTTACCTTCCTGCTCGAAGATGATTAT  8802
      |||||  :::::
Db  1615  -----Ilepecysersersersaphe  1622
```

Search completed: April 28, 2003, 13:38:05
Job time : 717 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: April 28, 2003, 11:02:51 ; Search time 96.5 Seconds
(without alignments)
6480.970 Million cell updates/sec

Title: US-10-087-013-1
Perfect score: 19435
Sequence: 1 atggggctctctgcaata.....atatatgatgaagattg 10628

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame.n2p.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US10087013/runat.28042003.102846/app.query.fasta_1.10823
-DB=Issued_Patents_AA -QPMF=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10087013.ecgn.1.159.etrna.28042003.102846.20293 -NCPU=6 -ICPU=3
-NO_XPPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-VARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	3501	18.0	3060	2	US-08-487-826B-14
2	2972.5	15.3	2710	2	US-08-568-459A-12
3	2972.5	15.3	2710	2	US-08-487-826B-12
4	2972.5	15.3	2710	4	US-09-210-288-12
5	1755	9.0	2182	2	US-08-487-826B-16
6	1032.5	5.3	700	2	US-08-568-459A-10
7	1032.5	5.3	700	2	US-08-487-826B-10
8	1032.5	5.3	700	4	US-09-210-288-10
9	649.5	3.3	921	2	US-08-568-459A-8
10	649.5	3.3	921	2	US-08-487-826B-8
11	649.5	3.3	921	4	US-09-210-288-8
12	618	3.2	1435	2	US-08-568-459A-4

13	618	3.2	1435	2	US-08-487-826B-4	Sequence 4, Appl1
14	618	3.2	1435	2	US-09-210-288-4	Sequence 4, Appl1
15	431	2.2	749	2	US-08-568-459A-6	Sequence 6, Appl1
16	431	2.2	749	2	US-08-487-826B-6	Sequence 6, Appl1
17	431	2.2	749	4	US-09-210-288-6	Sequence 6, Appl1
18	418	2.1	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
19	384.5	2.0	3135	1	US-08-323-170B-2	Sequence 2, Appl1
20	384.5	2.0	3135	4	US-08-954-441-2	Sequence 2, Appl1
21	383	2.0	3248	5	US-08-353-700-1	Sequence 1, Appl1
22	383	2.0	3248	5	PCT-US95-16216-1	Sequence 1, Appl1
23	365	1.9	411	2	US-08-568-459A-20	Sequence 20, Appl
24	365	1.9	411	2	US-08-487-826B-32	Sequence 32, Appl
25	365	1.9	411	4	US-09-210-288-20	Sequence 20, Appl
26	349.5	1.8	3596	4	US-09-134-001C-5080	Sequence 5080, Ap
27	343.5	1.8	1115	2	US-08-568-459A-2	Sequence 2, Appl1
28	343.5	1.8	1115	4	US-08-487-826B-2	Sequence 2, Appl1
29	343.5	1.8	1115	2	US-09-210-288-2	Sequence 2, Appl1
30	343.5	1.8	1115	6	5198347-6	Patent No. 5198347
31	286	1.5	2482	1	US-08-328-254-6	Sequence 6, Appl1
32	287.5	1.5	3418	2	US-08-603-753D-4	Sequence 4, Appl1
33	287.5	1.5	3418	4	US-09-099-753-4	Sequence 4, Appl1
34	287.5	1.5	3418	4	US-08-986-106-4	Sequence 4, Appl1
35	283	1.5	3418	2	US-08-639-501-2	Sequence 2, Appl1
36	283	1.5	3418	3	US-09-044-946-2	Sequence 2, Appl1
37	283	1.5	3418	3	US-09-044-908-2	Sequence 2, Appl1
38	280	1.4	3418	3	US-08-755-587-44	Sequence 44, Appl
39	279	1.4	362	2	US-08-568-459A-18	Sequence 18, Appl
40	279	1.4	362	2	US-08-487-826B-30	Sequence 30, Appl
41	279	1.4	362	4	US-09-210-288-18	Sequence 18, Appl
42	276.5	1.4	2285	4	US-09-308-375-2	Sequence 2, Appl1
43	276	1.4	411	2	US-08-568-459A-19	Sequence 19, Appl
44	276	1.4	411	2	US-08-487-826B-31	Sequence 31, Appl
45	276	1.4	411	4	US-09-210-288-19	Sequence 19, Appl

ALIGNMENTS

RESULT 1
US-08-487-826B-14
Sequence 14, Application US/08487826B
Patent No. 5993827

GENERAL INFORMATION:
APPLICANT: Sam, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSER: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487, 826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH21.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550

[illegible]

OY	2599	ATGTGACATCCATTGGTGAACATTTTAAACAAGCGATGATCACCACCTTAAGTGTAATTT	2658
Dd	967	Mctcshthrserashnleuglshaslnleaspval-----gliserVal	980
OY	2659	GTT---GATGATTAGTTAATAATTCCTTTTGGGGGATGTTCTTCTATGCACCAAAATAT	2715
Dd	981	ThrllysasnspLysalaserHisserLeuleuglyAspvalGlneubalaValarVthr	1000
OY	2716	GAACCAAAACAAGTAATNCGAATNGTATTAAAAGAAACAATPACTTAAGGCCCAAAGA	2755
Dd	1001	AspAlaalaglunllelllysAtgTYrLYsAspGlInashnle-----Gln	1016
OY	2776	GTACTGCACCCA-----AAACACACAGACAACATCTGTGCAGCTATAGCTTACAGT	2826
Dd	1017	LeuthrasprroilleinglnInLysaspnglnGlnlaImecCyArxgalValargryrer	1036
OY	2827	TTCGCAGATATAGTGTATTAATTCGAGCAAGATCTCTGGGAAGAAACGGTCACATG	2866
Dd	1037	PhehlaspIeuglyAspIlelleargglyrAspmetrPrAspGluUnpslyserSer	1056
OY	2887	GTAAGCGTGCAGAGACATTTTGGAAACGTGTTTTGGTAATATCATVAAGCACTCAAAAGC	2946
Dd	1057	ThrsapmetGluThrArgleuIllethrrValPheLYsAsnllelyseluShIsaspLy	1076
OY	2947	--AAAGCAAAATGATTAATATTAATGATGAT-----GCCCCCAAAATATTAAAAATG	2994
Dd	1077	IleLYsAspAnProLYsTYrThnglyAspGLUsErLYsPrOlaATyLYsALysleu	1096
OY	2995	AGCGAAATTTGGTGGGAAAGCTATATAGACCAAGTATGGGAAGCCATGAATGTGATTA	3054
Dd	1097	ArgAlasprrTPrgGluAlahsnlhghslnglnValTrpArgAlametylscys----	1114
OY	3055	AAATATTGAAGATTAATATGGGGCACCAATCAACAACAAGTAGTATTGGCGATATAGT	3114
Dd	1115	-----AlathrLYsglylIlelecyS-----	1121
OY	3115	GATCATACACCATGATGATATATATATCCACAACAAAATTAAATGATGCAGCATGGCCA	3174
Dd	1122	ProglmetProVALaspARYrIleProGlnArgLeuArgTRPmethrcGluTrpAla	1141
OY	3175	GAATGGTCTCAGGAGTGGCAAGAAAAGGATATGATATAGTGAAGAGAAGTGTAAAGG	3234
Dd	1142	GluTrpPYrCYsLYsAlaglnserGlnGluTrpAspLYsleuLYsLYsIleCYsAlaAsp	1161
OY	3235	TGTAAAGATTAAGTAATGTGTCAAGCGCTGNCGMAAGAGTGTACAGGTGTACGAG	3294
Dd	1162	CysmetSerLYsGLyspGLys----CysThr-----GlnglyAspValAspCyselys	1179
OY	3295	TGCACAGAACTTGTATGAATTAATATGATATATATAGATTTATGCAAGAACAAATGAT	3354
Dd	1180	CysLYsAlaAlacYsAspLYsTYrLYsGlnGlnIleGlnLYsTrpAsnGlnUlnTrpArg	1199
OY	3355	ATAATATCATGATTAATATCAAAAGAAATTCATGATGACACCAAAATGCTGTGTATAGT	3414
Dd	1200	LysIleSerAspLYsTYrAsnleuleuTYrLeuGlnAlaLYsThrThrSerThrasnPro	1219
OY	3415	GGTATTGAAGCTTCCAGTACTGCCAAAATCATATAGAC-----AGGAAATGTATWT	3465
Dd	1220	Gly-----ArgThrValLeuGlyAspAspAspProAspTYrGlnGlnmetVal:	1235
OY	3466	GAATTTTGTGCGAATTAATATCAACAAAATGCGGGCAAP-----AGTAAATAAA	3513
Dd	1236	AspPheleuinrProIleHlsLYsAlaserIleAlaAlaArgValLeuValLYsArgAla	1255
OY	3514	AGGTGACTACTGATGAAGAAGTGTGTCATTTGTGTATACACCAAGATCAAAATTTTGA	3573
Dd	1256	AlaGlySerProThrGluIleAlaAlaAlaAlaProIleThrProTySerThrAlaAla	1275
OY	3574	GCATATCCCATGATACAGGAAATTTTGATGATGTTCACATGCACAAAATAGATTTTGAT	3633
Dd	1276	GlyTYrIleHlsnglnIleGlyTYrGlyGlyCYsGelnglnGlnUlnThrclnPhesyGlu	1295

```
OY 3634 GAAAAA-----AGTGTGTAAGGATACGAAAAATATGCTTGA 3675
Db 1296 LysLysHisGlyAlaThrSerThrSerThrLysGlnSerGlyTyrThrPheLys 1315
OY 3676 GATAACACGACGACGATGATGCGTGTGCTTGAATGATGAAACGACAAAG 3735
Db 1316 GlnProProGlnTyrAlaThrAlaCysAspCysIleAsnArgSerGlnThrGlu--- 1334
OY 3736 GTACAGATTAACGAAAAAGCGAAGAAAAAGATGCGAAATGTAACAGTCAAT 3795
Db 1335 -----GlnProLysLysLysGlnGlnValGlnSerAlaCysLysIleValGln 1351
OY 3796 GATTAATTAAGAAAAAGATGGAAGAAACAGTGAAGATTCATCCAAAAAGAT 3855
Db 1352 LysIleLeuGlnGlyLysAsnGlyArgThrThrValGlyCysAsnProLysGlnSer 1371
OY 3856 AGTAATGATATCCCGATGGCAATGC---GGAATTAATTAATTAATTAATTAAT 3912
Db 1372 -----TyrProAspThrPaspCysLysAsnAsnIleAspIleSerHisAsp---Gly 1387
OY 3913 GTGTGATGCCCCCTAGACAGCAAAAGTTATGCTACATTTCTGGCAATGATTAAGAA 3972
Db 1388 AlAcysMetProProArgArgGlnLysLeuGlnSerLysLeuTyrIleAlaHisGlnSerGln 1407
OY 3973 ATAAAAAATTAACATCAACAGTAATTAATTAAGAAAGCTTTCAATTCAGACAGCA 4032
Db 1408 ThrGlnAsnIleLysThrPaspAspAsnLeuLysAspAlaPheIleLysThrAlaIala 1427
OY 4033 GAACATCTCTTCATGATGATTAATTAATTAAGTAAG---CATGCTGAAGAAATGAAC 4089
Db 1428 GlnThrPheLeuSerThrPglTyrLysSerLysAsnSperGlnAlaLysIleLeu 1447
OY 4090 GATTAAGATTAAGAAAGGCAAAATTCCTCCGCAATTTTGGATCCATGCTTCACACA 4149
Db 1448 AspArg-----GlyLeuIleProSerGlnPheLeuArgSerMetLeuTyrThr 1463
OY 4150 TTGGAGATTAATGAGATTTTATTTGACAGATATTCAAAGAGTATGCTGAGGGA 4209
Db 1464 PheGlyAspTyrAspLysIleCysLeuAsnThrAspIleSerLysGlnAsnSperVal 1483
OY 4210 ACTAACTAAAGACGAAATAGATTCCTCTTTCAAAATAGTGACCAAAATCTCCTAT 4269
Db 1484 AlaLysAlaLysAspLysIleGlyLysPhePheSerLysAspGlySerLysSerProSer 1503
OY 4270 GGAAGAAACGCCACAGATGCTGACAGACATAGTCATGATATGGAACCTATGCTA 4329
Db 1504 GlyLeuSerArgGlnGlnIleTyrPyrThrAsnGlyProGlnIleTyrPysGlyMetLeu 1523
OY 4330 TGTGCACTAGTAAATAATGGGCAAAAAAAGATGAT-----TTTACGAAATACATC 4380
Db 1524 CysAlaLeuThrLysTyrValThrAspThrAspAsnLysArgLysIleLysAsnSperTyr 1543
OY 4381 GGTTCACACACGCAATTAATTAAGTACAAA---AGCACCACTTTGAGGAATTTCCCAA 4437
Db 1544 SerTyrAspLysValAsnGlnIleSerGlnAsnGlnAsnProSerLeuGlnIlePheAlaIala 1563
OY 4438 CGACCCGAGTTTTCAGATGCTACACGAATGCTACGACGACTATGCTATACAGCAAA 4497
Db 1564 LysProGlnPheLeuArgTyrMetIleGlyTyrPyrGlyGlnGlnPheCysAlaGlnArgGln 1583
OY 4498 AATATTTGAAGATGTCAGGAAAAATGTAAGTCA---AATGACCAATTTGAGGATGAT 4554
Db 1584 LysLysGlnAsnIleIleLysAspAlaCysAsnGlnIleAsnSerThrGlnGlnCysAsn 1603
OY 4555 ACA-----GAATGTAATTAAGAAATGCGACGACTACGTTAATATATATGAAA---AAA 4602
Db 1604 AspAlaLysHisArgCysAsnGlnIleAlaCysArgAlaTyrGlnGlnTyrValGlnAsnLys 1623
OY 4603 AAAAAAGATGATTCACACAGATTAATTAATTAAGATGACCGACGACAAAAAGATTC 4662
Db 1624 LysLysGlnPheSerGlyGlnThrAsnAsnPheValLeuLysAlaAsnValGlnProGln 1643
OY 4663 GATAGACAAACATTTGTTAATGTTTACAGACTACTGACGAAAGATGCAACAGATTAC 4722
OY 1644 AspProGlnTyrLysGlyTyrGlnTyrLysAsp----- 1654
OY 4723 TTGAACAGAAATTTACTGCTAGTGTGTGCTGATTAAGCCCTGAAGCTCTGTGTACAA 4782
Db 1655 -----GlyValGlnPro-----IleGln 1660
OY 4783 AGAATATACATTTGTTAGAAAAACAGGCTTACTATGATGCCGACAAACATTTGGGTGC 4842
Db 1661 GlyAsnGlnTyrLeuLeuGlnLys-----CysAspAsnAsnLys---CysSerCys 1676
OY 4843 ACAAAATTTATGAATAATACGACCAAAATTAATTAACATTTGAGTAAGATTAAGTCAAA 4902
Db 1676 ----- 1676
OY 4903 GGATTAATTAAGAGGCAACACACAGTGTATTAAGTGCAAAACAAAGTCTAATTAAC 4962
Db 1676 ----- 1676
OY 4963 TACAACTACTGAAGAATTAAGTGAAGATGCTTTTCTCTCTGCTGACTACGTATA 5022
Db 1676 ----- 1676
OY 5023 TGTTCATGATGATGCAATTAATTAACAGATCCAGAAATTAAGATTAAGATTAAGGTTG 5082
Db 1677 -----MetAspLysAsnValLeuSerValSerProLysGlnLys----- 1689
OY 5083 CGAAAAAGATTAAGTGAAGTGGCGGCAACGAGAGGTCATTAATTTGGGTCATTAACAAA 5142
Db 1690 -----PropionylLysTyr----- 1694
OY 5143 GAAAAAAGAAAAAGAAAAAATTAATAAAGCTGCGATCGGCAAAATTTCTTAAAGGTC 5202
Db 1695 -----AlaHisLysTyrPro----- 1699
OY 5203 CCGCTGTAGTGTATGAATATAGTTTATGATTAATTAAGATTAATTAATTTCTAGTATT 5262
Db 1699 ----- 1699
OY 5263 GATTAATTTGAAGATGAAAAACAAAGACCGAGGAAAAATTTGAAGAAATATTAACAAA 5322
Db 1699 ----- 1699
OY 5323 AATGACATCATGTTGGCAAAAGAGTATGATGATCTACAGAAATCCCGTAGTACTCG 5382
Db 1699 ----- 1699
OY 5383 CGAAATTTTCTGGAACGAAATTAAGAAATGCTGTGAACGCAATGATATGCGGTAC 5442
Db 1699 ----- 1699
OY 5443 AAAGTGTAGGATGATGCAAAATAGTGAATAGTGAAGAAAGATGATGAAGATCTAATA 5502
Db 1699 ----- 1699
OY 5503 AATGTGTTCTGTACTTCATGATGATTAATCTATATGCGGAAAAATGCGATGAAGGT 5562
Db 1699 ----- 1699
OY 5563 ACTGCGTATCAGTTTCTTCGATGATGTTGCCGAATGGGTGAAGATTTTTCGAACATATA 5622
Db 1699 ----- 1699
OY 5623 GAAAGGAATTTGAGAAATTTGTAAGGGCGCTGAATGATTAATTAATTAATTAATTAAT 5682
Db 1699 ----- 1699
OY 5683 GATTAAGAAAGAAATGATCAGATGCGTGTACCAATTAATTAATTAATTAATTAATTAAT 5742
Db 1700 -----GluLysCys----- 1702
OY 5743 AAACGACAGTATGAAAAACAAATCAAAAAATATGTTGAGAAATTAAGACAAATAATATTCC 5802
```

[illegible][illegible]

QY 7501 AAATTATACCTTGTGCTACAAAGGAAAGTAAATAGTTGGACTAATAT 7560
Db 2231 -----
QY 7561 AATCTTCGATCTAGAAACCTTATGACCTGATTAATATATAGGAGAAACACCT 7620
Db 2239 -----
QY 7621 TGTGAAATAGAGAGAAATCGTTTAAAGTAGATTATGAAATGTTACAAAAT 7680
Db 2243 CysArgAsn-----
QY 7681 TCAAGTCTATCAGAGAGAAAGAGTATGTCTCTCCAGAGAGAACATATGTGC 7740
Db 2248 Ser-----
QY 7741 TTAAGAAATTTAGATGAATAATTAATGAAAGACTTAAGATAGTAATATCTCTAATA 7800
Db 2249 -----
QY 7801 ATGCTTCGTGCAACTGCAGGAAATAGAGAAATAGACATATATAAAACTTCAACTCA 7857
Db 2266 ArgValSerAlaAspSerLysSer-----
QY 7858 -----GAGACGGGTCGCCAATGAAATCAATATGTGATCTATGAAATATAGTTGCT 7911
Db 2279 GlyLeuGluAsnAlaCys-----
QY 7912 GATCTGGGTGACATAGTTAGAGAACAGATATGTTACAAATGGTGTACTTACCTGCC 7971
Db 2285 -----
QY 7972 GTAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8031
Db 2289 -----
QY 8032 AAAGTAGAATAATAATACACATGATGACAAAGCTTCTGCTGCTGGAGATCTAAT 8091
Db 2293 -----
QY 8092 ACAAAAGATATTTGGAAGCA-----ATGACGTGCAGAACCCAGAA 8133
Db 2294 ArgLysAspGluTrpValGluTrpPhe-----
QY 8134 GATGCAAAACTTTTGAAGAGAGAAATGATGATTTGAACGATACATTAATACAA 8193
Db 2313 AsnValAsnGlyGluAlaLysGlyLys-----
QY 8194 GATAGGTGACATAGAGAGATCCACCTGTGATGATTAATATACCTCAACGGTTTGA 8253
Db 2322 -----
QY 8254 TGGATGACTGAATGCTGGAATATATTTGTAAGCAGTATGAGAAATTTG 8313
Db 2329 LeuValLysArgTrpValGluTrpPhe-----
QY 8314 AAAAAATCATGTGACACTGTAAACATCTGACAGATGCAAGAAATGATTAATGAGAAAT 8373
Db 2345 LysHis-----
QY 8374 AAGTGTGACAGTGTAAACAGAGATGTCAGAAATATATAAATTTTGTCTTAATGAGAA 8433
Db 2356 GluLysSerProCys-----
QY 8434 TCTCTATTCGATATACATCAATAATTAATACAAAGATTTGTATGAGAACCAATATATACA 8493
Db 2369 -----
QY 8494 AAAATCTCTACTTATGATCATGTTCAAAATTTTGTACAAAAGTTGAAACCTTTAAAGT 8553
Db 2383 -----
QY 8554 GAATGTTCTGTGAGAGCTTTTCTGAATATCTCATGAACAGTAAGTGTGAATATAT 8613

Db 2386 Glu-----
QY 8614 AAATTATGAATAATGANGTCTTCCATATATACAGATATATGCTTTCGAGAAACACCA 8673
Db 2388 Lys-----
QY 8674 AAA-----
Db 2407 GlnIleThrAspAlaAsnAlaLysAsnValLysLysSerLysPheGlyAsnSer 2426
QY 8692 TCACAGTTACACTCTCTTCAGAAATCCATGATTAATGTCTACGATCAAAACAA 8751
Db 2427 CysGlyCysSerAlaSerAla-----
QY 8752 GATGATGTAGGAATTTCAAACTTTTACCTCTGCTGCAAGAAATGATTAATATAT 8811
Db 2439 AsnGly-----
QY 8812 CTGTATATTTGACGATACCTTGTCTTATAGTTCAATGATTAACAAAGTGTATG 8871
Db 2440 -----
QY 8872 ATTCCTCCAGAGAGACATTTATATGACAGACCTATCACTGATATATATAGAAA 8931
Db 2440 -----
QY 8932 GGTGATTAAGAAATTTTAAAAAACTTCTACTCTGCTTCAAGTCAAGACAATG 8991
Db 2440 -----
QY 8992 TTAGTCAAAAATTAATATCGAAGAGAGTGTGCTTTGAGCAATGAATATATAT 9051
Db 2440 -----
QY 9052 GCAGATTATTCGATATATTAATTAAGAACTGATATGATGACACTTATATCTGAAAA 9111
Db 2441 -----
QY 9112 ATTAATAAAATATTTGAACATCAATGAACCAACCAAAATGTAATAACATGCTGGAA 9171
Db 2457 Ile-----
QY 9172 AATATAGAGCTCAGATATGCGACGCTATGTATGTGATATAAATTTGCTACTCAAA 9231
Db 2457 -----
QY 9232 GTAACTATGATGAGATGCTGCAATTAACAAAGATGAGAAACTATCAGTTCTT 9291
Db 2458 -----
QY 9292 CGTGTAAATTTGATGAGGCAAGCAAGCATGTAAGGAAAAAGAACTGTAAGTATCA 9351
Db 2461 -----
QY 9352 TTAATAAATAATGT-----
Db 2469 -----
QY 9397 TCGAATTTATTAAGACAACTGATGTCAGATGATTAATTAATTAATTAATTAAT 9456
Db 2484 -----
QY 9457 ATATGTATAAATAATCAATGAAATATCAATATTAATTAATTAATTAATTAATTAAT 9516
Db 2498 AsnMetLeuTrpLysLysLysGluAsnVal-----
QY 9517 TCTTCAGGTATATAGACATAAACCATCTGAGAAAGAAAGTTCAGCATATTAATAATCA 9576
Db 2517 -----
QY 9577 AAGATTCTCAATGCGCTTTGAGATTAAATGATATTAATTAATTAATTAATTAATTAAT 9636

[illegible]

```

QY 10540 AATCTTGAGGATCTTATAGAGATGATTATATATGATGAGTGCTA 10599
      |||:::||||| || ||||| ||||| ||||| ||||| |||||
Db 2851 AsnAspLysAsnAspLeuTYrSerGlyIleAspLeuIleAsnAspAlaLeu--SerGlyA 2870
      ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 10600 AACCTATGATATATATGATGAGTATG 10628
      ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2870 snHsIleAspIleTYrAspGluMetLeu 2879
      ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
US-08-568-459A-12
: Sequence 12, Application US/08568459A
: Patent No. 5849306
: GENERAL INFORMATION:
: APPLICANT: Sim, Kim L.
: APPLICANT: Chitnis, Chetan
: APPLICANT: Miller, Louis H.
: APPLICANT: Peterson, David S.
: APPLICANT: Su, Xin-zhaun
: APPLICANT: Wellens, Thomas E.
: TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
: TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe Martens Olson & Bear
: STREET: 620 Newport Center Drive 16th Floor
: CITY: Newport Beach
: STATE: California
: COUNTRY: US
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/568,459A
: FILING DATE: 07-DEC-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelson, Ned
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER: NIH121.001CPI
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2710 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Plasmodium falciparum
US-08-568-459A-12

Alignment Scores:
Pred. NO: 1.08e-217 Length: 2710
Score: 2972.50 Matches: 865
Percent Similarity: 40.06% Conservative: 442
Best local Similarity: 26.51% Mismatches: 873
Query Match: 15.28% Indels: 1083
DB: 2 Gaps: 120

US-10-087-013-1 (1-10628) x US-08-568-459A-12 (1-2710)
QY 25 ATTATTAATAATGGGAATGACAGATCATCATCTTATGAGGAGATGCTAAAGCCCTATTATA 84
      ::: ||||| ||| :::: ||| :::
Db 6 LeuAlaLysMetGlyProLysGluAlaIaIaGlyGlyAspAspIleGluAsp----- 22
      85 AAAGAAAGTCACAAAGTGCAGAGAAATGTTTGGACGCTTATGCCCAAAAATATATAGACAT 144

```

```

Db 23 -----GluSerAlaLysHisMetPheAspArgIleGlyLysAspValTyrAsp 38
QY 145 CCATCAAAA---TATGCAAAAAGACATGTGGATTCGTGGAAGGATTTGACGAAAGCA 201
Db 39 LysValLysGluGluAlaLysGluArgLysGlyLeuGlnGlyArgLeuSerGluAla 58
QY 202 GAATTCGTGGTGGTCTCTTACGCGAGTAATAAGCATTAATTAATTCATTCATCA 261
Db 59 LysPheGluLysAsnGlnSerAspProGlnThrProGluAsp-----Pro 73
QY 262 TGTAAATTAGTCAATAGCAACATACATTAATTCAGGATATGATGATGATTTGAGACAT 321
Db 74 CysAspLeuAspHisLysTyrHisThrAsnVal-----ThrThrAsnValIleAsn 90
QY 322 CCTTCCATGGTATGAGAACAAACCGATTTGATGAGATGAGATGATCTGAAATG---GGA 378
Db 91 ProCysAlaAspArgSerAspValArgPheSerAspGluValArgLysGlnCysThrHis 110
QY 379 AATAAATAGCTATATTAATAAAGAAAAATGATGCTATACCTGTGCGCCACCTAGAGA 438
Db 111 AsnArgLleLysAspSerGlnGlnGlyAspAsnLysGlyAlaCysAlaProTyrArgArg 130
QY 439 CGACATATGTGTATATAAACTTGAAGCTCTAATGATATTAATCCCAAAATATTCAT 498
Db 131 LeuHisValCysAspGlnAsnLeuGlnGlnIleGlnProIleLysIleThrAsnThrHis 150
QY 499 GATTTATTTGGAAATGTACTACTTACAGCAAAATACGAAGGTAATGTAATTAATAT 558
Db 151 AsnLeuLeuValAspValArgLysMetAlaLysPheGlnGlyGlnSerIleThrGlnAsp 170
QY 559 CATCCA---CATAAAGAACT-----TCGAGCGTTGTACTCTCTTGA 600
Db 171 TyrProLysTyrGlnAlaThrLysArgLysAspSerProSerGlnIleCysThrMetLeuAla 190
QY 601 CGAAGTTTTCAGATATAGATATTTGAGAGAAATGATATGTTTAAACCAATGTC 660
Db 191 ArgSerPheAlaAspIleGlyAspIleValArgGlyArgAspLeuTyrLeuGlnLysnPro 210
QY 661 CATGAC-----AAATAGAAACGGGTCTCCGAGAGGTTTCAAGAAAAATA 705
Db 211 GlnGluLleLysGlnArgGlnGlnLeuGlnAsnLeuLysThrIlePheGlyLysIle 230
QY 706 CATGAT-----GGAATGGAAGATGAAATGAAATAAATGATTACAAATCCGATGATCT 756
Db 231 TyrGluLysLeuAsnGlyAlaGluAlaArgTyrGlyAsnAsp-----ProGlu----- 246
QY 757 GGAATTTATTAATTAAGAGACATGGTGAATGTGAATAGAAATTAAGTATGGGA 816
Db 247 -----PhePheLysLeuArgGlnAspTrpTrpThrAlaAsnArgGluThrValTyrPys 264
QY 817 GCTAATACATGATGATCATCATATTAATCTGATATTTTATGCAATCGAAAGTAATCA 876
Db 265 AlaIleThrCysAsnAla---TyrGlyAsnThrTyrPhe---HisAlaThrCysAsnArg 282
QY 877 CCATTAATTTCAATCTTAATTCGGCGCATTAACAGAAAGGATTCCTACCAATTTAGAT 936
Db 283 GlnGluArgThrLysGlyTyrCysArgCysAsnAspArgLysValProThrTyrPheAsp 302
QY 937 TATGTCCTCATATATTAATGTTGTTGAGCAATGGAAGAGGTTTGGCGAAAAAGA 996
Db 303 TyrValProGlnTyrLeuAlaGlyTrpPheGlnIleTyrPheLysPheCysArgLysLys 322
QY 997 AATATTAATTAAGAAAGGTCACAGACCTCGTCTGCT-----AATGCAAAAGAA 1044
Db 323 AsnLysLysIleLysAspValLysArgAsnCysArgLysLysAspLysGluAspLysAsp 342
QY 1045 CGCTTAATTTACTCATTAATGACATGATTTAGCAACAATAATTTGAAAAAGATAT 1104
Db 343 Arg---TyrCysSerArgAsnGlyTyrAspCysGlnLysThrLysArgAlaIleGlyLys 361
QY 1105 TTGCATTTGGATATATAGTCTAGCTGTTCCACTAAATGCAAAAGTTTGAAGTTGG 1164

```

```

Db 362 LeuArgTyrGlyLysGlnCysIleSerCysLeuTyrAlaCysAsnProTyrValAspTyr 381
QY 1165 TTAGGGAATCAACACAGAACATTTTAAAAAAGAAAAAATATGAAAAAGAAATACAA 1224
Db 382 IleAsnAsnGlnLysGlnGlnIlePheAspLysGlnLysLysTyrAspGlnIleLys 401
QY 1225 TCATAT-----TATGCAACGATMACAAATTTGTCAAT 1257
Db 402 LysTyrGluAsnGlyAlaSerGlyLysArgGlnLysArgAspAlaGlyTyrThr 421
QY 1258 AATATTAATTAAGGAA---TATATTAACAATTTTGAAGAACTTAAGCAACCAATAT 1314
Db 422 ThrThrAsnTyrAspGlyTyrGlnLysLysPheTyrGlnLeuAsnLysSerGlnTyr 441
QY 1315 GCACTAATGACACTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1362
Db 442 ArgThrValAspLysPheLeuGlnLysLeuSerAsnGlnGlnIleCysThrLysValLys 461
QY 1363 -----AAAGAGATTTACAGAGAAAGATATTTACTTTACTTAACAGTCT----- 1410
Db 462 AspGlnGluGlyGlyTyrIleAspPheLysAsnValAsnSerAspSerThrSerGlyAla 481
QY 1411 -----GATGACAAAGGATATTTATGCTTCGCAATATTTGCAAGTGTGT 1455
Db 482 SerGlyThrAsnValGlnSerGlnGlyTyrPheTyrArgSerLysTyrCysGlnProCys 501
QY 1456 CCCGACGCGGGGGTCAAAATGATGATGATTAATAATACACACAC---AAATCAGATTAATGAT 1512
Db 502 ProTyrCysGlyValLys-----LysValAsnAsnGlnGlyLysSerSerAsnGln 517
QY 1513 CGTGAACGTGTAAATTAAT-----GAAGCTATAAAGCTCCATGGCGT 1554
Db 518 TrpGluLysAsnAsnGlnLysCysLysSerGlyLysLeuTyrGlnProLysProAsp 537
QY 1555 GTGAACCTACTAATATACCTGTCCTTATAGCTGAATGAACAGAGTATTAACAA 1614
Db 538 LysGlnGlyThrThrIleThrIleLeuLysSerGlyLysGlyHisAspIleGlnGlu 557
QY 1615 AATTAAGAAATTTTGTAC-----ACCTCA 1641
Db 558 LysLeuAsnLysPheCysAspGluLysAsnGlyAspThrIleAsnSerGlyLysSerGly 577
QY 1642 ACTAATTAACAAAGTAATAAATTAAT-----CAAAATGGGAATGC 1680
Db 578 ThrGlyLysSerGlyLysGlnAsnSerGlyArgGlnGlnLeuTyrGlnGlnIleTyrPys 597
QY 1681 TATTAATAGATGAATAATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
Db 598 ---TyrLysGlyLysAsnValValLysValGlyHisAspGluAspAspGlnLysPyr 616
QY 1741 GATTAATCTTAAG-----ATPAATA----- 1758
Db 617 GluAsnValLysAsnAlaGlyLysLeuCysIleLeuLysAsnGlnLysLysAsnLysGlu 636
QY 1759 -----TCATTTCAATATTT 1773
Db 637 GlnGlyLysAsnThrSerGlnLysGluProAspGlnIleGlnLysThrPheAsnProPhe 656
QY 1774 TTGAATTTATGGGTTTCATATTAATTAATTAAGGATTAATTAATTAATTAATTAATTAAT 1833
Db 657 PheTyrTyrValAlaHisMetLeuLysAspSerIleHisTyrLysLysLeuGln 676
QY 1834 ACTTGATTA---AATATTAACAACACGCAATGCT---ATTGATGAATGTAACGAATATTCG 1887
Db 677 ArgCysLeuGlnAsnGlyAsnArgLleLysCysGlyAsnAsnLysCysAsnAsnAspCys 696
QY 1888 TATGTTTGAAGATGGGTTTAAACAAAAAGAAAGATGATATAGTATAAGAAAGCTG 1947
Db 697 GluCysPheLysArgTrpIleThrGlnLysLysAspGlnTyrGlyLysIleValGlnHis 716
QY 1948 TTCACAAAAAAGAAATATACAG----- 1971
Db 717 Phe---LysThrGlnAsnIleLysGlyArgGlyLysSerAspAsnThrAlaGluLeuIle 735

```


QY 1972 -----CAATCGTATTAGTAAATATTAAATCTTTTGAAGGTATTTT----- 2016
Db 736 ProheapsHisAspTyrValLeuGlnTyrAsnLeuGlnGlnGlnPheLeuLysGlyAsp 755
QY 2017 -----TTTAAAGTTATGATTAACCTTACAAAGATGAAAGCAAAATGCG 2058
Db 756 SerGlnAspAlaSerGlnGlnGlnSerGlnAsnSerLeuAspAlaGlnGlnGlnA----- 773
QY 2059 AAGAACTTATGAAATATATAAAGAAAGAAATAGATTTTCCATTTTGAAGAAATAT 2118
Db 774 -----GlnGlnLeuLysHisLeuArgGlnLeuLeuGlnGlnGlnGlnGlnGlnGln 789
QY 2119 AGGAC-----TATTAGAGATGCAATGAGACTTTGTTA 2154
Db 790 GlnGlnLysSerValGlnGlnGlnValThrGlnGlnLysAsnIleMetAspLysLeu 809
QY 2155 GATCACTTAAAGAACTGCGCATATCT----- 2184
Db 810 AsnTyrGlnLysAspGlnLysAlaAspLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 829
QY 2185 ---AAGACAAATATATACAAAGCAAGCATGTGAAACATCCATATGCAACAAACCCG 2241
Db 830 LysGlnLysGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 849
QY 2242 TGTGTTAACTCTGCGAGCAGCAAA---CCGACTAAATAATTAAGAAATAGCAAA 2298
Db 850 CysSerGlnLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 869
QY 2299 TACTTTAAAGAGTGCATGACGAGAAACGAAATCGTGTCTCTCAATTTGAAAGCA 2358
Db 870 MetHisHisLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 889
QY 2359 AAGGACACGAGAT 2418
Db 890 AspIleSerLeuAlaGlnPheLysAsnGlnArgAsnGlnSerThrLeuLysGlnGln 909
QY 2419 TGTAGATATATGATTAACATTTCTATCTGATCTGTTTTCATGACCAATGAT 2478
Db 910 CysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 929
QY 2479 GCGAAGACGACAGTGTAT 2538
Db 930 GlnLys---AspLysAspHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 948
QY 2539 GATCCGGAACACATGCGTAAAGATCAAGAGATGTTATGCTCTAGAACAGCAT 2598
Db 949 IleGlnGlnLysLysGlnThrSerThrLysAsnValPheLeuProProArgGlnHis 968
QY 2599 ATATGATCATCATTTTGAACATTTTCAAAAGAGATGATCAACCATTAATGATATAT 2658
Db 969 MetCysThrSerAsnLeuGlnAsnLeuAspVal-----GlySerVal 982
QY 2659 GTT---GATGATTTAGTATATATCTTTTGGGGAGTGTCTCTATGACAAATAT 2715
Db 983 ThrLysAsnAspLysAlaSerHisSerLeuGlnLysAspAlaGlnLeuAlaAlaLysThr 1002
QY 2716 GAAGCAAAACAGAT 2775
Db 1003 AspAlaAlaLeuIleLysArgTyrLysAspGlnAsnGln-----Gln 1018
QY 2776 GTAACGACCA-----AAACACCAACAATATCTGACAGTATACGTTACGT 2826
Db 1019 LeuThrAspProIleGlnGlnLysAspGlnGlnLysAspGlnGlnGlnGlnGlnGlnGln 1038
QY 2827 TTTGCAATATAGTAT 2886
Db 1039 PheAlaAspLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1058
QY 2887 GTAAGCGCAAGACATTTGAAATCTTTTGGTATATATATATATATATATATATATAT 2946
Db 1059 ThrAspMetGlnThrArgLeuIleThrValPheLysAsnIleLysGlnLysHisAspGln 1078

QY 2947 ---AAGGAAATGATTAAT 2994
Db 1079 IleLysAspAsnProLysTyrThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1098
QY 2995 AGGCAAAATATGCGGAGCAT 3054
Db 1099 ArgAlaAspThrProGlnLysAlaAsnArgHisGlnValThrArgAlaMetLysCys----- 1116
QY 3055 AAATATTTTGAAGATTAAT 3114
Db 1117 -----AlaThrLysGlnLysLysLysLysLysLysLysLysLysLysLysLys 1123
QY 3115 GATCATATACCATGATGAT 3174
Db 1124 ProGlnLysProValAspAspTyrIleProGlnArgGlnGlnGlnGlnGlnGlnGln 1143
QY 3175 GAATGATACGCAAGGCGCAAGAAAGGATATATATATATATATATATATATATATATAT 3234
Db 1144 GlnThrPyrCysLysAlaGlnSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1163
QY 3235 TGTAGGATATAGAT 3294
Db 1164 CysMetSerLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1181
QY 3295 TGCACAGAACCTTGTAT 3354
Db 1182 CysLysAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1201
QY 3355 AT 3414
Db 1202 LysIleSerAspLysTyrAsnLeuLeuLysLysLysLysLysLysLysLysLysLys 1221
QY 3415 GGTATGTAAGCTTCCAGTACTGCGCAAAATCATATATATATATATATATATATATAT 3465
Db 1222 Gln-----ArgThrValLeuGlnLysAspAspAspProAspPyrGlnGlnMetVal 1237
QY 3466 GAATTTTTCGCAAT 3513
Db 1238 AspPheLeuThrProIleHisLysLysAlaSerIleAlaAlaArgValLeuValLysArgAla 1257
QY 3514 AGTGTACTATGATGATTAAT 3573
Db 1258 AlaGlnSerProThrGlnIleAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1277
QY 3574 GCATATCTCCATGAT 3633
Db 1278 GlnTyrIleHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1297
QY 3634 GAAAAA-----AGTATGAT 3675
Db 1298 LysLysHisGlnLysAlaThrSerThrThrLysGlnLysAsnGlnGlnGlnGlnGlnGln 1317
QY 3676 GATTAACACGACGACAT 3735
Db 1318 GlnProProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1336
QY 3736 GATCAGATTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3795
Db 1337 -----GlnProLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1353
QY 3796 GATATATTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3855
Db 1354 LysIleLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1373
QY 3856 AGTATATGAT 3912
Db 1374 -----TyrProAspThrAspCysLysAsnAsnIleAspIleSerHisAsp---Gly 1389
QY 3913 GTGTATATGCCCCCTAGAACAAAGTATATATATATATATATATATATATATATATATATAT 3972
Db 1390 AlaCysMetProProAspArgGlnLysLysLysLysLysLysLysLysLysLysLys 1409
QY 3973 ATTAATAAAT 4032

```

Db 1410 ThrGlnAsnIleLysThrAspAsnLeuLysAspAlaPheIleLysThrAlaIala 1429
QY 4033 GAAACATTCCTTCATGATGATATATATAAAGTAG-----GATGGGAAGAAAGCACTC 4089
Db 1430 GluThrPheLeuSerIlePheIleLysThrAspSerGlnAlaLysIleLeu 1449
QY 4090 GATAAAGAAATTAAGAGCAAAATTCCTCCCGCATTTTGAGATCCATGCTTACACA 4149
Db 1450 AspArg-----GlyLeuIleProSerGlnPheLeuArgSerMetLeuIleThr 1465
QY 4150 TTGGAGATTATAGATTTTATTATGACACAGTATATCAAAAGCTCATGTGAGAGA 4209
Db 1466 PheGlyAspTyrAlaGlyAspIleCysLeuAsnThrAspIleSerLysGlnAsnAspVal 1485
QY 4210 AGTAAACTAAAGAGCAAAATGATCTCTTTCAAAAATGGTGACAAAATCCCTCAAT 4269
Db 1486 AlalysAlalysAspLysIleGlyLysPheSerLysAspGlySerLysSerProser 1505
QY 4270 GGAAGAAACAGCCCAAGATGCTGACAGACATAGCATGATATGGAGAGCTATGCTA 4329
Db 1506 GlyLeuSerAlaGlnGlnIlePheLysThrAsnGlyProGlnIleTyrLysGlyMetLeu 1525
QY 4330 TGTCACATGATAAAATTTGGGCAAAAAGATGAT-----TTTACCAGAAACTAC 4380
Db 1526 CysAlaLeuThrLysTyrValIleThrAspThrAspAsnLysArgLysIleLysAsnAspTyr 1545
QY 4381 GGTTCACACACGTCGAATTTAGTACAAA--AGCACACTTTGGAGAAATTTGCCAAA 4437
Db 1546 SerTyrAspLysValAlaGlnSerGlnAsnGlyAsnProSerLeuGlnIlePheAlaIala 1565
QY 4438 CGACCCAGTTTTCATGATGCTGATACCGATGATACGACACTATGCTATACACAGCA 4497
Db 1566 LysProIlePheLeuArgTyrMetIleGlnIleGlyLeuIlePheCysAlaGlnIle 1585
QY 4498 AATATTTGAAAGATGTGACGAGAAATGTAAGTCA--AATGACCAATTAAGTGTGAT 4554
Db 1586 LysLysGlnAsnIleIleLysAspAlaCysAsnGlnIleAsnSerThrIleGlnIleCysAsn 1605
QY 4555 ACA-----GATGTATATAGAAATGCGAGACTACGTTAATATATGAAA--AAA 4602
Db 1606 AspAlaLysIleHisArgCysAsnGlnAlaCysArgAlaTyrGlnIleLysValGlnIle 1625
QY 4603 AAAAAAGAGTGGATTCACAGATAATATTTACAGAGTACGACGACGACAAAAAGATTC 4662
Db 1626 LysLysGlnPheSerGlyIleThrAsnAsnPheValIleLysAlaAsnValGlnProGln 1645
QY 4663 GATAGACACACATGCTGATGTTACAGACTATCTGACAGCAATGCAACAGATTAC 4722
Db 1646 AspProGlnIleLysGlyLysIleLysAsp----- 1656
QY 4723 TTGAACAGAAATTTACTGCTAGTGTGATAGCCTGGAAGTGCCTCTGTGTACAA 4782
Db 1657 -----GlyValGlnPro-----IleGln 1662
QY 4783 ACAATATATACATTTGTTAGAAAAACAGCGCTTACTATGATGCGACAAACATTTGGTGC 4842
Db 1663 GlyAsnGlnTyrLeuLeuGlnLys-----CysAspAsnAsnLys--CysSerCys 1678
QY 4843 ACAAAATTTATTTGAAGATGACGACAAATATACTAATCATTTGAGATTAAGATGCAAA 4902
Db 1678 ----- 1678
QY 4903 GGATTTAGTAAGAGGCAACACAGGTGCTATTAAGTGCAAAAAGAGCTAATATAC 4962
Db 1678 ----- 1678
QY 4963 TACAATACTTGAAGAAATGACTGAAGATGCTTTTCTCTCTGCTGACTACGTATA 5022
Db 1678 ----- 1678
QY 5023 TGTTTTATGATGATGCAATTAATACAGATCCAGAGTTAAAGATGAAGATGGCTTG 5082
Db 5023 ----- 5082

```

```

Db 1679 -----MetAspGlyAsnValLeuSerValSerProLysGlnLys----- 1691
QY 5083 CGAAAAAGATTGATGGAAGTGGCGGACAGGAGGATACAATTTGGGTCAATACAGAA 5142
Db 1692 -----ProPheGlyLysTyr----- 1696
QY 5143 GAAAAAAGAAAAAGAAAAATTAATAAACGTGCGATCCGACAAATATCTATAGAGTC 5202
Db 1697 -----AlaHisLysTyrPro----- 1701
QY 5203 CCGCTTTAGTGTCTATGAATATAGTTTATTGATTTAAGATATTAATCTAGGTATT 5262
Db 1701 ----- 1701
QY 5263 GATTAATTTGGAAGATGAAAAACAAAAAGCCGAGAAAAATTTGAAGAAATATTAACAA 5322
Db 1701 ----- 1701
QY 5323 AATGCAACATCAGTTGGCAAGAAAGATAGTACTACAGAGAAATCCGATGACTGCG 5382
Db 1701 ----- 1701
QY 5383 CGAAAAATTTCTGGACGAAATAAGAAATGATGTGTGGAACCAATGATATCGGGTAC 5442
Db 1701 ----- 1701
QY 5443 AACGTGTAGGATGATGAAATAGTGAAATAGTGCAAGAGTGAAGATCTTAAA 5502
Db 1701 ----- 1701
QY 5503 AAATGTGTTCTGTACTTACCTTGATGATGATATATCTTATGGGAAAAATCCGATGAAGT 5562
Db 1701 ----- 1701
QY 5563 ACTGCTATCAGTTCTTCGATGTTGCCGAATGGGGTGAAGATTTTTCACAAATATA 5622
Db 1701 ----- 1701
QY 5623 GAAAAAGAAATGAGAAATTTGTAAGGGCGGTAAATGATATATCTTGTGTATATGAA 5682
Db 1701 ----- 1701
QY 5683 GATTAAGAAAAAGAAATGTACAGATGCGTGPACAAATATTAATAATTTATAGTAGTG 5742
Db 1702 -----GlnLysCys----- 1704
QY 5743 AAACACAGATATGAAAAACAATCAAAATATATGTGAGAAATAAAGCAAAATATATTC 5802
Db 1704 ----- 1704
QY 5803 GAGCATCTGTGCAAAAAGATGACAGAGACGCTCGGAATATTTAGACAAACAATTA 5862
Db 1704 ----- 1704
QY 5863 AAAATTTGTGAATAATTAAGTGAGATTGTGAATATATAGTATGAAGATGTGCACA 5922
Db 1705 -----AspCys----- 1706
QY 5923 CAGCGATTAAGTGTATAGTCAAAATATGCCGATCTTTAGCAGATGAACCAAA 5982
Db 1706 ----- 1706
QY 5983 GAAGTTGAAGAAAGTGTATATGTCAAGTCCGACGAGTCCACGACGTATGCAAGGAA 6042
Db 1707 -----TyrGlnGlyLys-----HisValProSerIleProPro----- 1717
QY 6043 ACACGCTACACAGGGTATCACTGATATCAAAAGCAGCAGCATCGAAGAAAGAGCGAAA 6102
Db 1717 ----- 1717
QY 6103 ACAGCGCGCTTACAAACAGCCGAAAAAGTGAATAATCTAACACAGAAATGCGAGCA 6162
Db 1718 ---ProProProProValGlnProGln----- 1725

```

[illegible]

OY	6907	CAATTTTGAACATGCTTCACAGAAATGGCGCAGAAGATTGTGAATAAAGGAGAACG	6966
Dd	2065	ProtyrPheargtgytleuglueglutrrgclvglnasPhenecysLysylsarglyshLys	2084
OY	6967	TTGTTAAAATTGAGGGCGGCGTGTAAGCAATATGAC-	7002
Dd	2085	LennlaeglInleuShLSGLucySLysValGlUGlnlaSngLyGlySerIArgGly	2104
OY	7003	-----TGTAAAT-----GGTAGTAAT	7017
Dd	2105	GlyllethrargIntyrserGIYAspSGlyUAlaCySaSngLUmetLeuProLysasn	2124
OY	7018	GACGGTAAGCACACAAGAA-----TGCGCAGAGCGCTGTCTAACTATADA	7062
Dd	2125	AspGIlythrValProaspLeugluLysProSerCySalalaLysProCysSerSetryIArg	2144
OY	7063	AATTTTATTAAAGCATGGAACATGCATTAATGAAGAACAAGAAAGTAAGTCAAAGAT	7122
Dd	2145	LysrrPIlegluserLySGlyLysGluPheglULysGlnglULysalatyGelUGln	2164
OY	7123	AAAGAT-----	7128
Dd	2165	LysaspLysCySaValaSngLyserrAnLysaSPasnLysaSPasnGlyPheCySglUthrLeuthr	2184
OY	7129	---GGCAAAAAAGTATAAGATATAT-----CCTTCTACTCGAAAGAGACATA	7170
Dd	2185	ThisSerSetrySalalaLysaspPheleuLysThrLeugLIProCySLysProasnaSnvaL	2204
OY	7171	GAGAAGCGCAACATGCTGCATGAAATATTTAAACATCAAAATTAAGAATTAATGTGCCAAT	7230
Dd	2205	GlUGlyLysThrIlePheaspPaspLysThrPheLys-----HisThr	2219
OY	7231	AAGGATGTCTCTGTATGACAAAAACCCTTCTCCACAACATCCAAAAACAACAACATCA	7290
Dd	2220	LysaspCySaSPProCySLeulysPheSerValaSncyLysLys-----	2234
OY	7231	CAATCATCCGATGCTATGATATGCCAGAAATGCGTGATTAATGTTCCTGAAGAATTAC	7350
Dd	2235	-----ASP	2235
OY	7351	AAGGTGAGTGTCCTGAACCTTCAAAAAAGCATCTATGATTCATCAAAAAAAATTACT	7410
Dd	2236	GlUCySaSP-----AnSerLysGlyThraSPcySaArgasnLys-----	2248
OY	7411	GAACCTAAATACCTATGAATGTGTGAGCAAAAAAGCAGCATATTAATTTATCTAAAGACA	7470
Dd	2249	-----AnSerIleasp-----AlethrSpIle	2256
OY	7471	GAATAATATATGATATTAATACCTTGGAAGAAAAATTTATACCTATTAAGCTTCAAAAGGA	7530
Dd	2257	GlUsnngLYaLaSPSerThrval-----LeuGlUmefarYalSer	2270
OY	7531	AAGGAAGTAAATTAATTAATGAGCTAATAATATCCCTGGAGCTCCAAGAAACCTTATGCA	7590
Dd	2271	AlaasPSerLysSerGIYPhesAnGlyasp-----	2280
OY	7591	CCTGATTAATATATATAGCAAGAAAGACCCCTGTGAAAT-----	7629
Dd	2281	-----GlyLeuGlUmSnlacYsaArgGlyAlaGlyIlePhegluclyLile	2295
OY	7630	AGAGAGAAAAATCGTTTTAAAGTAGATTAATGAATGGAATGTTACAAAAATTCAAAGTTC	7669
Dd	2236	ArgLysasp-----GlUtrPLycyS-----ArgsnValCySgLy	2307
OY	7690	TATCATGGAGAAAAAAGATATGTACACGCCCAAGAGAAACATATGTGCTTAAGGAT	7749
Dd	2308	ThrVal-----ValCySLysPro-----	2313
OY	7750	TTAGATCAAAATTAATTAAGAAAGACTTAAGATAGTAATTAATTCCTCAAAAATGTTGCT	7809
Dd	2313	-----	2313
OY	7810	CGAACTCAGCAAAATGAAGATATGACATATATAAAAAACCCTCACTCAGAGACGGTGC	7869

Db 790 GlnGluAlaSerValGlyGlyGlyValThrGlnGlnLysAsnIleMetAspLysLeuLeu 809
QY 2155 GATCACTTAAAGAAAGTGCACGATATGT----- 2184
Db 810 AsnTyrGlnLysAspGlnAlaAspLeuGlyLeuGlnIleHisGlnAspGlnGlnGln 829
QY 2185 ---AAAGACATATATACAAAGCAAGCATGTGAACATCCATTAATGCACAAACACCCG 2241
Db 830 LysGlnLysGlnLysAspGlnLysAsnGlnCysIleGlnGlnGlnLysAsnPheArgTyrAsnPro 849
QY 2242 TGTGTAAACCTCGTSGAGCAGCAGCA---CCACATAAAATTAAGAAATAGACACA 2298
Db 850 CysSerGlnLysGlnLysAsnLysArgTyrProValLeuAlaAsnLysValAlaTyrGln 869
QY 2299 TACTTAAAGAGTGCATACGAGGAGACAGAAATCGTCTGCTTCAATTAATGAAGA 2358
Db 870 MetHisLysLysAlaLysThrGlnLeuAlaSerArgAlaGlyArgSerAlaLeuArgGly 889
QY 2359 AAGCACACGAGTATATTAATTAACGTGGGCGTAGAGAAAGACATTCAAGCAATTTA 2418
Db 890 AspIleSerLeuAlaGlnPheLysAsnGlyAlaArgAsnGlySerThrLeuLysGlnIle 909
QY 2419 TGTAGATTAATGATTAACATTCCTTAATCGTATCTGTTTCAATGAGACATGAT 2478
Db 910 CysLysIleLeuAsnGlnAsnTyrSerAsnAspSerArgGlyAsnSerGlnLysProCysThr 929
QY 2479 GGCAGAGCGCAGGTGATGTATACAAACAGATTGTCGTAGGAACGTAAATGGGAAGT 2538
Db 930 GlyLys---AspGlnLysAspLysGlnGlyValArgMetArgIleGlyThrGlnTyrPserAsn 948
QY 2539 GATCCGGAACACATCGCTTAAGATCAGAAAGTATTAATTAATGCTTCAAGACACAT 2598
Db 949 IleGlnGlyLysLysGlnThrSerTyrLysAsnValPheLeuProArgLysGlnLys 968
QY 2599 ATATGTACATCCATTTGCAACATTTCACAAACGATGATCACCCATTAATGATATAT 2658
Db 969 MetCysThrSerAsnLeuLysAsnLeuAspVal-----GlySerVal 982
QY 2659 GGT---GATGATTTAGTAAATTCCTTTGGGGAGTCTCTCTATACACAAATAT 2715
Db 983 ThrLysAsnAspLysAlaSerHisSerLeuGlnLysAspValGlnLeuAlaLysThr 1002
QY 2716 GAAGCAACAGATTAATGCAATGTAATAAGAAAGAAATTAACCTTAAGGCCCAAGAA 2775
Db 1003 AspAlaAlaGlnIleLysLysArgTyrLysAspGlnAsnAsnIle-----Gln 1018
QY 2776 GTAACCTGACCCA-----AAACACGACACTATCTGCGATACGTTACAGT 2826
Db 1019 LeuThrAspProIleGlnGlnLysAspGlnLysAlaMetCysArgAlaValArgTyrSer 1038
2827 TTTCGAGATATAGTATTAATTCGAGAGAGATCTCTGGGAAAGAAACGCTGACATG 2886
Db 1039 PheAlaAspLeuGlyAspIleLysGlnLysArgAspMetThrPaspGlnAspLysSer 1058
QY 2887 GTAAGCTGCAAGACATTTGGAAACGTTTTGGTAAATATACAAAGCTCATAAAGC 2946
Db 1059 ThrAspMetGlnThrArgLeuLysPheLysAsnIleLysGlnLysHisAspGly 1078
QY 2947 ---AAAGAAATGATAATTAATATGATAT-----GCCCCAAATATTAATATG 2994
Db 1079 IleLysAspAsnProLysTyrThrGlnLysAspLysLysProAlaTyrLysLysLeu 1098
QY 2995 AGGGAATTTGGTGGAGAGTAAATAGACCAAGATGGAAGCAATGATGATATA 3054
Db 1099 ArgAlaAspTyrPyrGlnLysAsnArgHisGlnValThrPaspAlaMetLysCys----- 1116
QY 3055 AAATATTTGAAGATTAATCGGACACCAACATCAACAAAGTACTTAATGGCGATATAGT 3114
Db 1117 -----AlaThrLysGlnIleLysCys----- 1123
QY 3115 GATCATACACATTTGATGATTTATATCCACAAATTAAGATGATGACCGAATGGCA 3174
Db 1124 ProGlyMetProValAspAspTyrIleProGlnArgLeuArgTyrMetThrGlnTyrAla 1143

QY 3175 GAATGCTACTCGAAGGTGCAGAAAAGAGATATGATTAAGTTGAAGAGAGATGAAGAG 3234
Db 1144 GlnTyrPyrCysLysAlaGlnSerGlnGlnIleLysAspLysLysLysIleCysAlaAsp 1163
QY 3235 TGTAAAGATTAAGATTAAGTCAAGGTGTCAGAAAGAGATGTCACAGTTGTCAGAG 3294
Db 1164 CysMetSerLysGlnLysAspLysLys---CysThr---GlnGlyAspValAspCysGlyLys 1181
QY 3295 TGCACAGAACTGTGAATGATATATATTAATTAAGATTAAGTGAAGAAAGCAATGCAAT 3354
Db 1182 CysLysAlaAlaCysAspLysTyrLysGlnGlnIleGlnLysThrPaspGlnGlnThrArg 1201
QY 3355 ATAAATGAGATTAATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 3414
Db 1202 LysIleSerAspLysTyrAsnLeuLysLeuGlnAlaLysThrThrSerThrAsnPro 1221
QY 3415 GGTATTTGAAGCTTCAGTACTGCCAAAATCATATATAGAC-----AGGAATGTTAT 3465
Db 1222 Gly-----ArgThrValLeuLysAspAspProAspTyrGlnGlnMetVal 1237
QY 3466 GAATTTTGTGGAAATTTATACCAACAAATGTGGCAAA-----AGTAATMAA 3513
Db 1238 AspPheLeuThrProIleHisLysLysSerIleAlaAlaArgValLeuValLysArgAla 1257
QY 3514 AGTGTACTAGTGAAGAGTGTGCTCATTTGCTACTAACACACAGTATGAAAATGTTGCA 3573
Db 1258 AlaGlySerProThrGlnIleAlaAlaAlaAlaAlaProIleThrProTyrSerThrAlaAla 1277
QY 3574 GCATCTCTCATGATATACAGAAATTTGATGATGTCAGTGCACAAATGAGTTTGTGAT 3633
Db 1278 GlyTyrIleHisGlnGlnIleGlyTyrGlnLysGlnGlnGlnGlnGlnPheCysGln 1297
QY 3634 GAAAA-----AGTATGCTAAGATTAAGCAAAAAATATGCTTTAGA 3675
Db 1298 LysLysHisGlnLysAlaThrSerThrSerThrLysGlnLysAsnLysGlnTyrThrPheLys 1317
QY 3676 GATAAACCAAGACATGATGTCGTGCTGTTGAAGATGATGGAACCGACAGCAAG 3735
Db 1318 GlnProProGlnTyrAlaThrAlaThrAlaCysAspCysLysAsnArgSerGlnThrGln 1336
QY 3736 GTACAGATTAATAAGCAAAAAAGCGGAAAGAAAGATACGGAATGTAAACAGTGAAT 3795
Db 1337 -----GlnProLysLysLysGlnLysAsnValGlnSerAlaCysLysIleValGln 1353
QY 3796 GATATCTTAAGAAAGCAATGATGGAAGAAACAGTGAAGATGTCATCCAAATGAAGAT 3855
Db 1354 LysIleLeuGlnGlnLysAsnGlnLysArgThrThrValGlnLysAsnProLysGlnSer 1373
QY 3856 AGTATGATATCCCGATTTGGCAATGC--GGAATATTAATTAATTAATGGAACCTCGT 3912
Db 1374 -----TyrProAspTyrPaspCysLysAsnAsnIleAspIleSerHisAsp---Gly 1389
QY 3913 GTGTATGTCGCCCTGAAACAAAGATTAATGCTGATATTTCTGGCAATGTAATGA 3972
Db 1390 AlaCysMetProProArgAlaGlnLysLysCysLeuTyrTyrIleAlaHisGlnSerGln 1409
QY 3973 ATAAAAATTAACATCAACAGTAAATTAATTAAGAGCTTCAATCAATCTGACAGAGA 4032
Db 1410 ThrGlnAsnLysThrAspAsnLeuLysAspAlaPheIleLysThrAlaAlaAla 1429
QY 4033 GAAACATTTCTTCATGATTAATTAATTAAGTAAAG--GATGCTGAAGCAATGAACTC 4089
Db 1430 GlnThrPheLeuSerTyrPyrGlnTyrLysSerLysAsnAspSerGlnAlaLysIleLeu 1449
QY 4090 GATTAAGAAATTAATAAGCAAAATTCCTCCGCAATTTTGAAGTCACTGTTTACACA 4149
Db 1450 AspArg-----GlyLeuIleProSerGlnPheLeuArgSerMetMetTyrThr 1465
QY 4150 TTTCGAGATTAATGAGATTTTATTTGACACAGATATTAAGAAAGTCAATGTCAGAGA 4209
Db 1466 PheGlyAspTyrArgAspIleCysLeuAsnThrAspIleSerLysLysGlnAsnAspVal 1485

Qy	7063	AAATTATTAAGAAAGCGAAAAACCTAATATGTAAGACAAAGAAAAGTTCAAAAAGAT	7122
Db	2145	LYSTRIPLLEIUSERISGLYLSGLIUPHEGLIULSGINGIULYSLATATYGLINGIN	2164
Qy	7123	AAAGAT-----	7122
Db	2165	LYSASPGLYSCYVALASNGLYSERASNULYSHISAPSNLYPHECYSGLUTHRIEUTH	2184
Qy	7129	----GGCAAAAATATATAGGATTAT-----	7170
Db	2185	ThrSerSerLysAlaIysAspPheLeuLysThrLeuGlyProCysLysProAsnVal	2204
Qy	7171	GAGAAAGCAACATGTGCTCAGCAAAATATTTAAACATGAAATTAAGAATTATGTGGCAAT	7230
Db	2205	GLUGLYLysThrIlePheAspAspAspLysThrPheLys-----HisThr	2219
Qy	7231	AAGGATTTGTTCTTGATGACAAAACCTTCCTCACACACTACAAAACAAACACACATCA	7290
Db	2220	LysAspCysAspProCysLeuLysPheSerValAsnCysLysLys-----	2234
Qy	7291	CAATCATCGAGTCAATGATGATGCCAGATACCGCTGATATATGTCTCTAAGAAATTATAC	7350
Db	2235	-----Asp	2235
Qy	7351	AAGGTGAGTGTCTCTGCACTTTCAAAAAGGATCTATGATTCATACAAAAAATTTACT	7410
Db	2236	GLUCYAsp-----AsnSerLysGlyThrAspCysArgAsnLys-----	2248
Qy	7411	GAACTTAAATACCTATGAAATTTGTCGTAAGAAGCAGCATATTTATTTATCTTAAACAGCA	7470
Db	2249	-----AsnSerIleAsp-----AlaThrAspIle	2256
Qy	7471	GAAATTAATATGATATTTACCTTGAAAGAAAAATTTATACCTATGAGTCTACAAAGGA	7530
Db	2257	GUaSnGlyValAspSerThrVal-----LeuGluMetArgValSer	2270
Qy	7531	AAGGAAGTAAATATAGTGGACATTAATATACCTTGCGATCTCAAGAAACCTTATGCA	7590
Db	2271	AlaAspSerLysSerGlyPheAsnGlyAsp-----	2280
Qy	7591	CCTGATTAATATATAGGAAGAGAAACCCTTGTGAAT-----	7629
Db	2281	-----GlyLeuGluAsnAlaCysArgGlyIleIlePheGluGlyIle	2295
Qy	7630	AGAGAAGAAATCGTTTAAAGTAGATATATGATGATGAATGTTACAAAATTCAAAGTC	7689
Db	2296	ArgLysAsp-----GluThrLysCys--ArgAsnValCysGly	2307
Qy	7690	TATCAGAGAAAAAAGATATGTACTCTCCACAGAAGAACATATGTGTTAAGAAAT	7749
Db	2308	TyrVal-----ValCysLysPro-----	2313
Qy	7750	TTAAGTGAATTAATTAATTGAAGACTTAAGATAGTAAATTATCTCTAAAAATGGTTCGT	7809
Db	2313	-----	2313
Qy	7810	CGAACTGACGAATGAGGAATAGACATATATAAAAACTTCACTCAGAGAACGGGTGC	7869
Db	2314	-----GluAsnValAsnGlyLeu-----	2319
Qy	7870	GCAATGATCCAAATATGTCAITACTATGAATATATAGTTTCGTGATCTGGGTGACATAGTT	7929
Db	2320	-----Ala	2320
Qy	7930	AGAGACAGACATATATGTAAGATATGGTGTACTTACCTACCTCCCGTAGAAATTAATATATAT	7989
Db	2321	LysGlyLysHisIleIleGlnIleArgAlaLeuValLys-----ArgThrVal	2336
Qy	7990	AAGTTTGGATATACATATATGGAATAAGCAATATAAATAAAGTAGAATTAATATAC	8049
Db	2337	GluTyrPhePheGluAspLysValAsnLysIleLysHisLysIleSerHisArgIleLysAsn	2356

```

QY 8050 AACGAT-----GTACAAAGCTTCTGCTGCTGGGATGCTAATAGAAA 8097
Db 2357 GlyIuILserProCysIleIysAsnCysValGluLysTyrValAspGluLysAspGly 2376
QY 8098 GATATTGGAAGCAATGACGTGCAAGACACCGAAGATGCAAAACTTTTACAAAAGA 8157
Db 2377 Glu---TrrLysGluIleThrGluAtpheLysAsp----- 2387
QY 8158 AGAATGATGATTTGACGCAATACATTAAATCAAGATAGTGTGACATTAAGACGAT 8217
Db 2388 -----GlnTyrLysAsnAspAsnSerAspAsp 2396
QY 8218 CCACCTGTGATGATATATATACCTCAACGTTTCGATGATGACCTGAATGGTGTGATAT 8277
Db 2397 AspAsnValArgSerPheLeuGluThrLeuIleProGluIleThrAspAlaAsn----- 2414
QY 8278 TATTGTAAGCACTGATGAGCAATGGAATAATTTAAAAATCAGTGTGATCTGTATAA 8337
Db 2415 ---AlaLysAsnLysValIleLysLeuSerLysPheGlyAsnSerCysGly---CysSer 2432
QY 8338 ACATCTGACAGATGCAAGATGATTTGATGAAATTAAGTGTGACAGTGTAAACGAGA 8397
Db 2433 AlaSer-----AlaAsnGluGlnAsnLysAsn----- 2441
QY 8398 TGTCAAGATATATAAATTTGTTCTTAATGGAATCTCTATTCGATATACATCAAT 8457
Db 2442 ---GlyLyuIrrLysAspAlaIle-----AspCysMetLeuLys 2453
QY 8458 AATATCAAGAAATTTGATGACAAACAAATATACAAAAATCTCTACTTATGATCATGTT 8517
Db 2454 LysLeuLysAsp-----LysIleGlyLysGluLysGly 2465
QY 8518 CAAATTTTGTACAAAGTTGAAACTTTTAAAGTGAATGTTCTGTGTGAGACCTTTTCT 8577
Db 2466 HIsHIs-----GlnThrSerAspThrGluCysSerAspThrProGlnPro 2480
QY 8578 GAATATCTTCATGAACAAAGTATGTTGATTAATTAATTAATTAATTAATGATGTTCT 8637
Db 2481 GlnThrLeuGluLysp-----GluThrLeuAspAspIleGluThrGluGluAlaLys 2498
QY 8638 TCCAAATATACGAACATATGCTTTCGAAGAAACCAAAAGTTATTAAGAA----- 8688
Db 2499 LysAsnMetMetProLysIleCysGluAsnValLeuLysThrAlaGlnGlnGluLyspGlu 2518
QY 8689 ---GCTTGCAGTTGTACACTACCTTCTAAGATTCATTTGATTAATTTGCTTACCGATCAA 8745
Db 2519 GlyLysCys-----ValProAlaGluAsnSerGluGluProAlaAlaThrAspSer 2535
Db 8746 AACAAAGAT 8754
Db 2536 GlyLysGlu 2538

```

```

COMPUTER READABLE FORM:
MEDION TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fullier, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1RWV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-12

Alignment Scores:
Pred. No.: 1,086-217 Length: 2710
Score: 2972.50 Matches: 865
Percent Similarity: 40.06% Conservative: 442
Best Local Similarity: 26.51% Mismatches: 873
Query Match: 15.28% Indels: 1083
DB: 4 Gaps: 120

US-10-087-013-1 (1-10628) x US-09-210-288-12 (1-2710)
QY 25 ATTATATAAATGGGAAATGACATCATGATTAAGAGAGATGCTAAAGCCCTATTATA 84
Db 6 LeuAlaLysMetCylProLysGluAlaIleGlyLysAspIleGluLysp----- 22
QY 85 AAGAAGTCACAAAGTCAAGAAATGTTTGAACGTTATGCCAAATTAAGACAT 144
Db 23 -----GlnSerAlaLysHisMetPheAspArgIleGlyLysAspValTyrAsp 38
QY 145 CCATCAAAA---TATGCAAAAGACATGTGATTCCTGAAAGGGGATTTGAGAAACA 201
Db 39 LysValLysGluGluAlaLysGluArgGlyLysGlyLeuGlnLysArgLeuSerGluAla 58
QY 202 GAATTTGCTGTGTGCTTCTACGCCAGTAATTAAGCATTAATTAATTAATTCATATCA 261
Db 59 LysPheGluLysAsnGlnLysSerProGlnThrProGluLysp-----Pro 73
QY 262 TGTAAATTAATGATCAAGACATCACTACTAATTTAAGCTATGATGATGATTTGAGACAT 321
Db 74 CysAspLeuAspHisLysTyrHisThrAsnVal-----ThrThrAsnValIleAsn 90
QY 322 CCTTGCATGTAGAGACAAACCGATTTGATGAAGATGAAGATCTGAATGT---GGA 378
Db 91 ProCysAlaAspArgSerAspValArgPheSerAspGluTyrGlyLysIleCysThrHis 110
QY 379 AATAAATACGTAAATTAATAAAGAAATAATGATGCTTAAGCCTGTGGCCACCTAGACAGA 438
Db 111 AsnArgIleLysAspSerGlnGlnGlnLysAspAsnLysGlyAlaCysAlaProTyrArgGly 130
QY 439 CGACATATGTGATTAATAAAGCTTGAAGCTTAATAAGATTAATAATCCCAAAATATCAT 498
Db 131 LeuHisValCysAspGlnAsnLeuGlnIleGluProIleLysIleThrAsnThrHis 150
QY 499 GATTTATGGAATGATGATTAAGCAAAATAGACAAATGATCAATCAATTAAT 558
Db 151 AsnLeuLeuValAspValCysMetAlaAlaLysPheGluGlyGlnSerIleThrGlnAsp 170

```

[illegible]

```
Db 870 MethHisHisLysAlaLysThrGlnLeuAlaSerArgAlaGlyArgSerAlaLeuArgGly 889
QY 2359 AAGGCAACAGAGATATATATAAAGCTGGGAGAGAGAAAGACATTCACAGCAATTTA 2418
Db 890 AspIleSerLeuAlaGlnPheLysAsnGlyArgAsnGlySerThrLeuLysGlyGlnIle 909
QY 2419 TGTAGATATATGATTAACATTCATTCGTAATCTGTTTTCATTAATGGACCATGTGAT 2478
Db 910 CysLysIleAsnGlnLysnTrpSerAsnAspSerArgGlyAsnSerGlyProCysThr 929
QY 2479 GCGAAAGGCAAGGTGATGATACAAACAGATTTGCTGAGGAAGTGAAGTGAAGTGC 2538
Db 930 GlyLys---AspGlyAspHisGlyGlyValArgMetArgIleGlyThrGlnTrpSerAsn 948
QY 2539 GATCCGGAACACATCGCTAAGATACAGAAAGATGTTATGATGCTCCTGACAGACAGAT 2598
Db 949 IleGlnGlyLysLysGlnTrpSerTrpLysAsnValPheLeuProProArgArgGlnHis 968
QY 2599 ATATGTACATCCAAATTTGGAACATTTACAAACGAGATGATCACCACCTTAATGTAATAT 2658
Db 969 MetCysThrSerAsnLeuGlnAsnLeuAspVal-----GlySerVal 982
QY 2659 GTT---GATGATTTATATATATTCCTTTTGGGGAGTGTCTTATCAGCAAAATAT 2715
Db 983 ThrLysAsnAspLysAlaSerHisSerLeuGlyAspValGlnLeuAlaAlaLysThr 1002
QY 2716 GAAGCAAAACAGATATACGATGTATTAAGAAAGAAATACCTAAAGGCGCCCAAGAA 2775
Db 1003 AspAlaAlaGlnIleLeuLysArgTrpLysAspGlnAsnHisIle-----Gln 1018
QY 2776 GTAAGTGAACCA-----AAACACAGACAACTATGCTGACATACGATACAGT 2826
Db 1019 LeuThrAspProIleGlnGlnLysAspGlnLysAlaIleMetCysArgAlaValArgTrpSer 1038
QY 2827 TTGSCGATATAGTGATATATATTCGAGAGAGATCTCGGAGAAAGAAAGCGTGACAT 2886
Db 1039 PheAlaAspLeuGlyAspIleLeuArgGlyArgAspMetTrpAspLysSerSer 1058
QY 2887 GTAAGCTGCAAGACATATGGAACCTGTTTGTGATATATACATATGATCAGTCAAGAGC 2946
Db 1059 ThrAspMetGlnThrArgLeuIleThrValPheLysAsnIleLysGlnLysHisAspGly 1078
QY 2947 ---AAAGAAATGATATATATATGATGAT-----GCCCCCAATATTTAAATG 2994
Db 1079 IleLysAspAsnProLysTrpThrGlyAspGlnSerLysLysProAlaTrpLysLeu 1098
QY 2995 AGGAAATATGCTGGAGAGTATATAGCCAAAGTATGGGAGAACCATGAAATGTGATATA 3054
Db 1099 ArgAlaAspTrpTrpLysAlaAsnArgHisGlnValTrpArgAlaMetLysCys----- 1116
QY 3055 AATATATTGAAGATTAATCGGACACATCAACAACAAGTATGTTATTCGGATATAGT 3114
Db 1117 -----AlaThrLysGlyLleIleCys----- 1123
QY 3115 GATCATACACCATGATGATTTATATCCCAAAAATTAAGATGATGACCAATGGGCA 3174
Db 1124 ProGlyMetProValAspAspTrpIleProGlnArgLeuArgTrpMetThrGlnTrpAla 1143
QY 3175 GAATGTACTGCAAGGTGCGAAAGAGCTATGATATAGTATGAGAGAGAGAGTAAAGAG 3234
Db 1144 GlnTrpTrpCysLysAlaGlnSerGlnGlnTrpAspLysLeuLysIleCysAlaAsp 1163
QY 3235 TGTAAAGATAGATATGCTCAAGCTGTACAAAGAGAGTGTACAGGTGTACAGCA 3294
Db 1164 CysMetSerLysGlyAspGlyLys---CysThr---GlnGlyAspValAspCysGlyLys 1181
QY 3295 TGCAACAGACCTGTATATGATATATATGATATATAGATATGAGAAAGACATGAGAT 3354
Db 1182 CysLysAlaIleLysAspLysTrpLysGlnGlnIleGlnLysTrpAsnGlnTrpArg 1201
QY 3355 ATATATACATTAATTAACAAGATTAAGATGAGACCAAGCCAAATGCTGTATATATAGT 3414
Db 1202 LysIleSerAspLysTrpAsnLeuLeuTrpLeuGlnAlaLysThrThrSerThrAsnPro 1221
QY 3415 GGTATGAAAGCTTCCAGTCTCGCAAAAATTCATATAGAC-----AGCAATGTATAT 3465
Db 1222 Gly-----ArgThrValLeuGlyAspAspAspProAspTrpGlnGlnMetVal 1237
QY 3466 GAATTTTGTGCGAATTAATACCAACAAATGCTGGCAAA-----AGTAAATAA 3513
Db 1238 AspPheLeuThrProIleHisLysAlaSerIleAlaIleArgValLeuValLysArgAla 1257
QY 3514 AGTGTACTAGATGATGAAAGTGTGATGCTGATGCTGATACACACACGATGATGAAATGTGGA 3573
Db 1258 AlaGlySerProThrGlnIleAlaAlaAlaProIleThrProTrpSerThrAlaAla 1277
QY 3574 GCATATCTCATGATACAGAAATTTGATGATGCTGACGTCAACCAAAATGATTTGTGAT 3633
Db 1278 GlyTrpIleHisGlnGlnIleGlyTrpGlyGlyGlyGlnGlnIleThrGlnPheCysGly 1297
QY 3634 GAAAA-----AGTATGTTAGGATTAACGAAATAATGCTTTAGA 3675
Db 1298 LysLysHisGlyAlaThrSerThrSerThrThrLysGlnLysGlnLysGlnTrpThrPheLys 1317
QY 3676 GATTAACACAGACCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3735
Db 1318 GlnProProGlnGlnLysArgLysAlaCysAspCysIleAsnArgSerGlnThrGln--- 1336
QY 3736 GTACAGATTAATAAAGCAAAAGCAAGAAAGCAAGATGCAAGATGCAAGATGCAAGAT 3795
Db 1337 -----GlnProLysLysLysGlnGlnLysAsnValGlnSerAlaCysLysIleValGln 1353
QY 3796 GATATCTTAAAGAAAGCAAGTGAAGAAAGCAAGTGAAGATGCTGCAATGCAAAAGAT 3855
Db 1354 LysIleLeuGlnGlnLysAsnGlyArgThrThrValGlnLysAsnProLysGlnSer 1373
QY 3856 AGTAATGATATCCGATGCGCAATGC-----GGAATATATATATGATGCAAGCTCTCT 3912
Db 1374 -----TrpProAspTrpAspCysLysAsnAsnIleAspIleSerHisAsp---Gly 1389
QY 3913 GTCGTATGCCCCCTGAGACAAAGTATGCGTATGCTGCAATGCTGCAATGATATGAA 3972
Db 1390 AlaCysMetProProArgArgGlnLysLeuGlyLeuTrpThrIleHisGlnSerGln 1409
QY 3973 ATTAATAATTAATCAACAGTATTAATAAAGAGCTTTCATCAATCTGCACAGCA 4032
Db 1410 ThrGlnAsnIleLysThrAspAsnLeuLysAspAlaPheIleLysThrAlaAlaAla 1429
QY 4033 GAAACATCTTCATGATATTAATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 4089
Db 1430 GlnThrPheLeuSerTrpGlnTrpLysSerLysAsnAspSerGlnAlaLysIleLeu 1449
QY 4090 GATTAAGATTAATAAAGCAAAATTCCTCCGCAATTTTGGAGATCAGTGTCTACACA 4149
Db 1450 AspArg-----GlyLeuIleProSerGlnPheLeuArgSerMetThrTrpThr 1465
QY 4150 TTGCGATATATGACATTTTATTTTGGACAGATATATCAAAAGCTCATGCTGAGGGA 4209
Db 1466 PheGlyAspTrpArgAspIleCysLeuAsnThrAspIleSerLysGlnAsnAspVal 1485
QY 4210 AGTAACTAAAGACGAATAGATCTCTTTCAAAAAGTGTACCAAAAATGCTCTAT 4269
Db 1486 AlaLysAlaLysAspLysIleGlyLysPhePheSerLysAspLysSerProSer 1505
QY 4270 GCAAAACACGCGCAAAAGTGTGACAGACAAATGATCATGATATGAGAGATATGATCA 4329
Db 1506 GlyLeuSerArgGlnGlnTrpTrpLysTrpAsnGlnProGlnIleTrpLysGlyMetLeu 1525
QY 4330 TGTGCACTAGTAAATAATGCGGCAAAAAGATGAT-----TTTACGAAAACATAC 4380
Db 1526 CysAlaLeuThrLysTrpValThrAspThrAspAsnLysArgLysIleLysAsnAspTrp 1545
QY 4381 GCTTACACAGACGTAATTAATGACGAAA-----AGACACACTTGGAGAAATTTGGCAAA 4437
Db 1546 SerTrpAspLysValAsnGlnSerGlnAsnGlyAsnProSerLeuGlnGlnPheAlaAla 1565
```

```
QY 4438 CGACCCAGTCTTTTACGATGGCTAACCGAATGGTACGACACTATTCCTATACACGACAA 4497
Db 1566 LysProGlnPheLeuArgTyrMetIleGluTyrGluLeuPheCysAlaGluValGln 1585
QY 4498 AATATATTGAAGATGTGCAGGAAAAATGTAAGTCA--AATGACCAATGAAGTGTGAT 4554
Db 1586 LysLysGluAsnIleLeuLysAspAlaCysAsnGlnLeuLeuSerThrGlnGlnCysAsn 1605
QY 4555 ACA-----GAATGATATAGAAATGGAGACACTTAAATATATGAAA---AAA 4602
Db 1606 AspAlaLysHisArgCysAsnGlnAlaCysArgAlaTyrGlnGluValGlnAsnLys 1625
QY 4603 AAAAAAGAGTGGATTCACACAGATTAATATTACAGAGATGACGACAAAAAAGATTC 4662
Db 1626 LysLysGluPheSerGlyGlnThrAsnAsnPheValLeuLysAlaAsnValGlnProGln 1645
QY 4663 GATAGACAAACACATGTGTGTAATGCTTACAGACTATACGAAATGCAACAGATTAC 4722
Db 1646 AspProGlnLysLysGlyTyrGluTyrLysAsp----- 1656
QY 4723 TTGACAGAGAAATTTACTGCTAGTGTGATAGCTGGAAGTGCCTGTGTGTACAA 4782
Db 1657 -----GlyValGlnPro-----IleGln 1662
QY 4783 AGAATATACATTTGTTAGAAAAACAGCTTACTATGATCCGACAAACATTTGTGGTGC 4842
Db 1663 GlyAsnGluTyrLeuLeuGlnLys-----CysAspAsnAsnLys--CysSerCys 1678
QY 4843 ACAAAATTTATGAAATATGACGACAAATATACTAATTTGCGATGAATATAGTCCAAA 4902
Db 1678 ----- 1678
QY 4903 GGATTAGTAAAGAGGACAAACACAGTGTCTATTAGTGGCAAAACAAAGTCTATATAC 4962
Db 1678 ----- 1678
QY 4963 TACAATACTGAAAGATTGACTGAAGATGTCCTTTTCTTCCTCCTGACTACGTATA 5022
Db 1678 ----- 1678
QY 5023 TGTTCATGCAATGGATGCAATATTATACAGATCCAGAGTTAAAGATGAAATGGGTTG 5082
Db 1679 -----MetAspLysAsnValLeuSerValSerProLysGluLys----- 1691
QY 5083 CGAAAAAGATTGATGGAAGTGGCGCACGGAAGGTTACATTTGGTCAATACTACAAA 5142
Db 1692 -----ProPheGlyLysTyr----- 1696
QY 5143 GAAAAAAGAAAAAGAGAAAAATTAACACGTCGATGGCGCAAAATATTCTTATGAGTTC 5202
Db 1697 -----AlaHisLysTyrPro----- 1701
QY 5203 CCGCCTTGTAGTCTATGAATATAGTTTATGATTAAAGATTAATTTCTAGTATT 5262
Db 1701 ----- 1701
QY 5263 GATTAATTGGAGATGAAAAACAAAAGACCGAGAAATTTGAGAAAATATTTAACAAA 5322
Db 1701 ----- 1701
QY 5323 AATGACATCAGTTGGCAAAAGAGATAGTACTACAGAAATCCGGTAGTACTGCG 5382
Db 1701 ----- 1701
QY 5383 CGAAAAATTTTCTGGAACGAAAAATAGGAATGTGTGGAACGCAATGATATGCGGTTAC 5442
Db 1701 ----- 1701
QY 5443 AAACGTGTAGGATGATGAAATAGTGAATAGTGCAGAGAGTATGATGAAGATCTAAA 5502
Db 1701 ----- 1701

QY 5503 AATATGTTCTGTACTTTCACATGATGATTATTCCTATGCGGAAAAATCGGATGAAGT 5562
Db 1701 ----- 1701
QY 5563 ACTGGTATCAGTTTCTTGATGATGTTTGGCCAAATGGGGAAGATTTTGCACAACTAAA 5622
Db 1701 ----- 1701
QY 5623 GAAAAAGATTGGAGAAATTTGGTAGGGCGGTATGATTATTAATTACTTGTGTATATGAA 5682
Db 1701 ----- 1701
QY 5683 GATTAAGAGAAATGTACAGATCGGTACACATATTAATAAATTTATTATGAGTGG 5742
Db 1702 -----GluLysCys----- 1704
QY 5743 AAACCAAGTATGAAAAACAAATCAAAAAATATGTGTGAGATTAAGACAAAAATATTTC 5802
Db 1704 ----- 1704
QY 5803 GAGCATCTGTGGCAAAAGATGACAGAGAGCGCTGCGAATATTAGACAAACATTAATA 5862
Db 1704 ----- 1704
QY 5863 AAAATTTGTGAATAAATAAGTGAAGATGTGATATATAAGTATGAAAGATGTGCACA 5922
Db 1705 -----AspCys----- 1706
QY 5923 CAGCATTAACATGATGTTAATAGTCAAAAAATATGCGCCGATCATTAGACATGAACCAA 5982
Db 1706 ----- 1706
QY 5983 GAAGTTGAAGAAAGTGAATTGTCAGTGGCCAGAGTGCACACGTATACAGAGAA 6042
Db 1707 -----TyrGlnGlyLys-----HisValProSerIleProGln----- 1717
QY 6043 ACACGTCACACGCGGTATCATCTATATCAAAACGACGCGATCGAAAAAAGACGAAA 6102
Db 1717 ----- 1717
QY 6103 ACAGCGCGCTTACAAAAACGACCGCAAAAAAGTGAATACTAACACAGAAATGCGAGCA 6162
Db 1718 -----ProProProProValGlnProGln----- 1725
QY 6163 CAACACGAAACCCGACGACGACACAAACAAACAGAAAGAACATCAACGACAGACA 6222
Db 1726 -----ProGluLysLysTyr----- 1730
QY 6223 ACAGAAATCTGACGTGGCACACAGTGAAGCCATTCCTTCGAATAAACAGATAGCAGG 6282
Db 1731 ValThrValAspValCysSerIleValLysThrLeuPheLys-----AspThrAsn 1747
QY 6283 GGTGAATAGAGGGTGTATATCCAAAGATGAGCAAAAT-----CCTAAATGGGT 6336
Db 1748 AsnPheSerAspAlaCysGlyLeuLys--TyrGlyLysThrAlaProSerSerTyrLys 1766
QY 6337 TGTATT-----GTAGTAAGTCAATAAGAAAT 6363
Db 1767 CysIleProSerAspThrLysSerGlyAlaGlyAlaThrThrLysLysSerGlySerAsp 1786
QY 6364 GAAATGCAATATGATGCTCTCTAGAGAAAAAATATGATTAATAATATATACATAT 6423
Db 1787 SerGlySerIleCysIleProArgArgArgLeuLysValGlyLysLeuGln--- 1805
QY 6424 TTAATATGAACT-----GAAATATAGCTGAC 6453
Db 1806 ---GluThrPalaThrAlaLeuProGlnGlyGluGlyAlaAlaProSerHisSerArgAla 1824
QY 6454 AATGATTAATAAGAGGCTTTTATTAATGTCAGCAATAGAACTAAATTTTGTGTTA 6513
Db 1825 AspAspLeuArgAsnAlaPheIleGlnSerAlaAlaIleGluThrPhePheLeuTyrPasp 1844
QY 6514 AATATATATATTAATA-----AATCTCGACGACGAAAAATGAA----- 6549
```


Db 1845 ArgGlyIysSerIleuLysLysProGlnGlyAspLysSerGlnGlnAlaLeuSerGlnLeu 1864
QY 6550 -----TTGCAAAATGGCAACA 6564
Db 1865 ThrSerThrIysSerAspAspGluAspProProAspLysLysLeuGlnAsnGlyLys 1884
QY 6565 ATTCACAGATGAATTTAAAGCAATATGATATACATATGCGATATTAAGATATGTTT 6624
Db 1885 IleProProAspPheLeuAlaGluMetPheIleThrLeuGlyAspIleLeu 1904
QY 6625 TTTCGAACGTATATTTGATATGATTAATAAATTAACGTACCAATAGTATACAAACC 6684
Db 1905 ValHisGlyIleLysIleThrSerAspSerGlyAsnThrAsnGlySerAsnAsnAsnIle 1924
QY 6685 ATTCCTCAATGAATAATATAAGAAAAACAGATTAATAAAGATTAACGATAA 6744
Db 1925 ValLeuGlnAlaSerGlyAsnLysGluAspMetGlnLysIleGlnGlyLysIleGln 1944
QY 6745 ATA----- 6747
Db 1945 IleLeuProLysAsnGlyIleThrProLeuValProLysSerSerAlaGlnThrProAsp 1964
QY 6748 ----TTTGGAGAAATAAATAAATTTATTTGGAGCAATGATATGATTAACCTAT 6804
Db 1965 LysTrpTrpAsnGlnHisAlaGlnSerIleTrpLysGlyMetIleCysAlaLeuThrTrp 1984
QY 6805 ----CATCTCAGACAGAAAAAGAAAAAGAAAAAATAATAGA 6840
Db 1985 ThrGluLysAsnProAspTrpIleSerAlaArgGlyAspGluAsnLysIleGluLysAspAsp 2004
QY 6841 GATATATAC----- 6849
Db 2005 GluValLysGluLysPhePheGlySerThrAlaAspLysHisGlyThrAlaSerThrPro 2024
QY 6850 ----CAGTACAAAT--GACATGACCAAACTG-- 6873
Db 2025 ThrGlyThrTrpLysTrpGlnTrpGlnTrpGluLysValLysLeuGluAspTrpIleSerGly 2044
QY 6874 ----ACGCTTCCTCCCTGAGAGAGTTTGTAAAAAGGCC 6906
Db 2045 AlaLysThrProSerAlaSerSerAspTrpProLeuLeuSerAspPheValIleuArgPro 2064
QY 6907 CAATTTTGGAGATGCTTCAACAGATGGCAGAGAAATTTGTATTAAGAGAGAGAAACAG 6966
Db 2065 ProGlyPheArgLysLeuGlnGlyIleGlnAsnProCysLysLysAspLysHisLys 2084
QY 6967 TTGTATAATTTGAGCGCGGCTGTAAAGCAATATGAC-- 7002
Db 2085 LeuAlaGlnIleLysHisGlyLysValGluGlnLysnGlyGlySerArgArgGly 2104
QY 7003 ----TGTAAAT-----GGTAGTAAT 7017
Db 2105 GlyIleThrArgIleLysSerGlyAspGlyGluAlaCysAsnGlnMetLeuProLysAsn 2124
QY 7018 GACCGTAAACACACAGAA-----TGTSCAGACGCGCTGTACATCAATCAAA 7062
Db 2125 AspIleLysValProAspLysGluLysProSerCysAlaLysProCysSerSerTrpArg 2144
QY 7063 AATTTTAAAGAGAGTGAAGATGATTAAGAAACAGAAAGAAAGTCAAAAGAT 7122
Db 2145 LysTrpIleGlnSerLysGlyLysGluPheGlnLysGlnGlnLysAlaTrpGluGlnGln 2164
QY 7123 AAAGAT----- 7128
Db 2165 LysAspLysCysValAsnGlySerAsnLysHisAspAsnGlnLysCysGluThrLeuThr 2184
QY 7129 ----GGCAAAAGTAAAGATAT-----CCTTCTACTGAAGACACATA 7170
Db 2185 ThrSerGlyLysAlaLysAspPheLeuLysThrLeuGlyLysProCysLysProAsnVal 2204
QY 7171 GAGAAAGCAACATGTGTCATGATATTTAAACATGAATTAAGAAATATATGCGCAAT 7230

Db 2205 GluGlyLysThrIlePheAspAspAspLysThrPheLys-----HisThr 2219
QY 7231 AAGGATTTGTTCTGTATGCAAAAACCTTCTTCAACATACCAAAAACACAAACATCA 7290
Db 2220 LysAspCysAspProCysLysLysPheSerValAsnCysLysLys----- 2234
QY 7291 CAATCATCGATGCTAATGATATGATCCAGAAATCGCTGATATATGTTCTCTCAAGATTTAAC 7350
Db 2235 -----Asp 2235
QY 7351 AAGTGAGTGCTCTGTAACCTTCAAAAAGGATCTATGATCATCAAAAAAATTACT 7410
Db 2236 GluCysAsp-----AsnSerLysGlyThrAspCysArgAsnLys----- 2248
QY 7411 GAACCTAAATACCTATGAATGTGTAGAGAAACAGACATATTTATTTATTAAGAGCA 7470
Db 2249 -----AsnSerIleAsp-----AlaThrAspIle 2256
QY 7471 GAAATATATGATGATATTCCTTGAAGAGAAATTTATACCTATGAGCTTACAAAGAA 7530
Db 2257 GluAsnGlyValAspSerThrVal-----LeuGluMetArgValSer 2270
QY 7531 AAGGAAAGTAAATAATGTTGACATAATATATATCCCTTGCGATCCTAAGAAACCTTATGCA 7590
Db 2271 AlaAspSerLysSerGlyPheAsnGlyAsp----- 2280
QY 7591 CCTGATTAATATATATAGCAAGAAACCCCTTGCAAAAT----- 7629
Db 2281 -----GlyLeuGluAsnAlaCysArgGlyAlaGlyIlePheGluGlyIle 2295
QY 7630 AGAGAGAAATATGCTTTAAAGTAGATTAATGAATGGAATGTTACAAAATTCAGAACTTC 7689
Db 2296 ArgLysAsp-----GluTrpLysCys--ArgAsnValCysGly 2307
QY 7690 TATCAGAGAGAAAAAGAGTAGTGATGCTCAAGAGAGAACATATGCTCTTAAGAAAT 7749
Db 2308 TyrVal-----ValCysLysPro----- 2313
QY 7750 TTAGATGAATTAATAATTAAGAAAGCTTAAGATATATATCTCTTAATAATGTTTCGT 7809
Db 2313 ----- 2313
QY 7810 CGAAGTCACAGAAATGAGAAATAGACATATATAAAAACCTTCAACCTCAGAAACGGCTGC 7869
Db 2314 -----GluAsnValAsnGlyLys----- 2319
QY 7870 GCAATGAATCCAAATATGATGATGATGATGAATATATGTTGCTGATCGGTGACATAGT 7929
Db 2320 -----Ala 2320
QY 7930 AGAGCAACAGATATGTTACGAATTTGTTACTTACCTCCCGTAGAAATTAATATAT 7989
Db 2321 LysGlyLysHisIleIleGlnIleArgAlaLeuValLys-----ArgTrpVal 2336
QY 7990 AAGCTTTTGAATACATATATGCAAAATGAGAAATTAATAATTAAGCTAGAAATTAATAC 8049
Db 2337 GluTrpPhePheGlnLysArgLysLysIleLysHisLysIleSerHisArgIleLysAsn 2356
QY 8050 AACGAT-----GACAAACAGCTTCTGCTCTGCTGGGAGTGTCTAATAGAAA 8097
Db 2357 GlyGluLysLeuSerProCysIleLysAsnCysValGluLysTrpValAspGlnLysArgLys 2376
QY 8098 GATATTTGGAAGCAATGACGTGCAAAAGCAACAGAAATGCAAAACCTTTTAAGAAAGCA 8157
Db 2377 Glu--TrpLysGluIleThrGluArgPheLysAsp----- 2387
QY 8158 AGAATGATGATTTGAACGCAATACATTAATACAGATAAAGTGTGACATTAAGGACAT 8217
Db 2388 -----GlnTrpLysAsnAspAsnSerAspAsp 2396
QY 8218 CCAAGCTGTGATGATTAATATACCTCAACAGGTTTCGATGATGATGATGATGATGAT 8277
Db 2397 AspAsnValArgSerPheLeuGluThrLeuIleProGlnIleThrAspAlaAsn----- 2414

QY 8278 TATGTAAAGACATGAGAGAAATTTGAAAAATTTAAAAATCATGACTGATAA 8337
 Db 2415 ---AlaLysAsnLysValIleLysLysSerLysHegLysAsnSerCysGly---CysSer 2432
 QY 8338 ACATCTGCACATGCAAGATGATTAATGATTAATAAATAGTGTGAACAGTGTAAACGAGA 8397
 Db 2433 Alaser-----AlaAsnLysLysAsn----- 2441
 QY 8398 TGTCAAGATATATAAATTTGTTTAAATGAAATCTGATCATATACATCAAT 8457
 Db 2442 ---GlyGluTyrLysAspAlaIle-----AspCysMetLeuLys 2453
 QY 8458 AAATACAAAGATGTGTGAACAAATATATACAAAAATCTCTACTATGATCATGTT 8517
 Db 2454 LysLeuLysAsp-----LysIleGlyLysGluLysLys 2465
 QY 8518 CAAATTTTGTACAAAAGTTGAAAACCTTTTAAAGTAAAGTCTGTGTGAGACCTTTCT 8577
 Db 2466 HisHis-----GluThrSerAspThrGlnCysSerAspThrProGlnPro 2480
 QY 8578 GAATATCTTCATGAAACAGTAAGTGTGTAATTAATTAATGAAATGAGTTC 8637
 Db 2481 GlnThrLeuAspAsp-----GluThrLeuAspAspIleGluThrGluAlaLys 2498
 QY 8638 TCCAAATATAGAACATATGCTTTCGAAGAACACCAAAAGTTATTAAGAA----- 8688
 Db 2499 LysAsnMetLeuProLysIleCysGluAsnValLeuLysThrAlaGlnGlnGluAspGlu 2518
 QY 8689 ---GCTTGCAGTTGTACACTACCTCTTAAGATTCATGATTAATGTTGCTACCGATCA 8745
 Db 2519 GlyGlyCys-----ValProAlaGluAsnSerGluGluProAlaAlaThrAspSer 2535
 QY 8746 AACAAAGAT 8754
 Db 2536 GlyLysGlu 2538
 RESULT 5
 US-08-487-826B-16
 ; Sequence 16, Application US/08487826B
 ; Patent No. 5993827
 ; GENERAL INFORMATION:
 ; APPLICANT: Sim, Kim L.
 ; APPLICANT: Chintis, Chetan
 ; APPLICANT: Miller, Louis H.
 ; APPLICANT: Peterson, David S.
 ; APPLICANT: Su, Xin-zhuan
 ; APPLICANT: Wellens, Thomas E.
 ; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 ; NUMBER OF SEQUENCES: 45
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbe Martens Olson & Bear
 ; STREET: 620 Newport Center Drive 16th Floor
 ; CITY: Newport Beach
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/487,826B
 ; FILING DATE: 10-SEP-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Israelson, Ned
 ; REGISTRATION NUMBER: 29,655
 ; REFERENCE/DOCKET NUMBER: NIH121.001CP1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176
 ; INFORMATION FOR SEQ ID NO: 16:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2182 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: 1linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: N-terminal
 ; US-08-487-826B-16
 Alignment Scores:
 Pred. No.: 4.13e-125 Length: 2182
 Score: 1755.00 Matches: 760
 Percent Similarity: 29.37% Conservative: 337
 Best Local Similarity: 20.35% Mismatches: 678
 Query Match: 9.02% Indels: 1960
 DB: 2 Gaps: 125
 US-10-087-013-1 (1-10628) x US-08-487-826B-16 (1-2182)
 QY 37 GGAATGACAGCATCATCATGAGGAGAGATGCTAAAGCCCTATATATAAAGAAAGTCAC 96
 Db 10 GlySerGlyGlySerSer---SerGlyLysGlyLysLysAspThrSerGluTyrIleLys 28
 QY 97 AAAAT---GCAGAAATGTTTGGACGT-----TATGCCAAAATATA 138
 Db 29 ValSerAspAlaLysAspLeuLeuAspArgValGlyGluLysValTyrGluGluLysVal 48
 QY 139 AGACATCCATCAAAATATATGCAAAAGACATGCTGGATTCGTAAGGAGGATTCACGAA 198
 Db 49 LysAsnGlyAsp-----AlaLysLysTyrIleGluAlaLeuLysGlyLysLeuAsnThr 66
 QY 199 GCAGATTCGTCGTGTCCTCTTACGCCACTAAATAGCATTAATTAATTCATAT 258
 Db 67 AlaAsnGlyArgSerSerGluThrAlaSerSerIleGlu----- 79
 QY 259 CCATGTAATTTAGATCAATTAAGACATCACTAATTTAGGATGATGATGCAAT----- 312
 Db 80 ThrCysThrLeu---ValLysGluTyr-----TyrGluArgValAsnGlyAsp 94
 QY 313 ---TTGACATCCCTGCCAT-----GGTAGACAAACAAACGATTTGATGAA 357
 Db 95 GlyLysArgHisProCysArgLysAspAlaLysAsnGluAspValAsnArgHisSerAsp 114
 QY 358 GATGAAGAATCTGAATGTGA---AATAAATACGTAATTAATAAAGAAATGATGCT 414
 Db 115 ThrLeuGlyGlyGlnCysThrTyrAsnArgIleLysAspSerGlnGlnLysAspAsnLys 134
 QY 415 ATA---GCTGTGGCCACCTAGAGACAGATATGATGATTAATAAAGTGAAGTCTGA 471
 Db 135 ValGlyAlaCysAlaArgProTyrArgArgLeuHisLeuLysAspTyrAsnLeuGluSerIle 154
 QY 472 AATGATATAATATCCCAAAATATTCATGATTTATGGAAATGCTACTAGTACGCAAA 531
 Db 155 Asp-----ThrThrSerThrThrHisLysLeuLeuLeuGluValAlaCysMetAlaLys 172
 QY 532 TACGAGTGAATCAATTTGTAATTAATCAAT-----CCATATAAGAACT----- 576
 Db 173 TyrGluLysAsnSerIle---AsnThrHisTyrThrGlnHisGlnArgThrAsnGluAsp 191
 QY 577 ---TCAGAGCCTGTACTGCTCTGACGAGATTTGACAGATTTAGATTAATTTGA 630
 Db 192 SerAlaSerGlnLeuLysThrValLeuAlaArgSerPheAlaAspIleLysPleVal 211
 QY 631 AGAGGAATATATGTTT-----AAACCAATCTCCATGACAAAGTA 672
 Db 212 ArgGlyLysAspLeuTyrLeuGlyTyrAspAsnLysGluLysGluLysLysLeu 231
 QY 673 GAAACGGCTCCGAGAGGTTTCAAGAAATATACATGATGAGATGCAATGCAATGATAA 732
 Db 673 GAAACGGCTCCGAGAGGTTTCAAGAAATATACATGATGAGATGCAATGCAATGATAA 732

OY 2743 AAAGAAAGAAATACCTAAAGGCCCCCAAGAAAGTAAGTACCCAAAACAGCAACT 2802
|||||
Db 824 lysglugllythrValthrGlValthrGlValthrGlValthrGlValthr--- 842
OY 2803 ATGTGTGACGCTATACGTTTACAGTTTCCAGATATAGTGTATTAATTCGAGAGAGAT 2862
842 ----- 842
OY 2863 CTCTGGAGAAAGAAAGCGTGACATGTAAAGCTGCAAGACATTTGAAACGTCTTTTGGT 2922
842 ----- 842
OY 2923 AATATACATAGTCACTCAAAAGCAAGAAAGAAATATTAATTAATGATGATGCCCCAAA 2982
842 ----- 842
OY 2983 TATTTAAATTTGAGGAAATTTGTGGGAAGCTAATAGACCAAAAGTATGGAAGCCATG 3042
842 ----- 842
OY 3043 AAATGTGATTAATAATTTTGAAGATTAATCGGGACCAATCAACACAAAGATTTAT 3102
842 ----- 842
OY 3103 TCGGATATAGTATCATACACCATTTGATGATTTATCCCAAAAATTAAGATGGAG 3162
843 ----- 843
OY 3163 ACCGAATGGCAGAAATGTACTGCAAGGTGCAAGAAAGAGATATGATTAAGTGAAGAG 3222
|||||
Db 844 ThrGlu----- 847
OY 3223 AAGGTGACGAGTGTAAAGATTAAGATTAATGCTCAAGGCTGTACGAAGAGAGTGTACA 3282
|||
Db 848 GlyValLysProCys----- 852
OY 3283 GGTTGTACGAAGTGCACAGAGCTTGTATGATTAATGATTAATGATTAATGATTAATGAAA 3342
853 ----- 858
OY 3343 GAACATGGAATTAATATCAGATTAATACAAAGAAATTAACATGACACAGCAAAATGTCT 3402
858 ----- 858
OY 3403 GTTAGTAATAGTGTATGTAAGCTTCCAGTACTGCCAAAAATCATATACAGAGAAATGTT 3462
858 ----- 858
OY 3463 ATTGAATTTTGTGGAATTAATACCAACAAATGTGCGCAAAAGTAATAAAGTGTACT 3522
858 ----- 858
OY 3523 AGTGATGAAGTGTCTCATTTGTACTAACACACAGTATGAAAAATGTGGACATATCTC 3582
858 ----- 858
OY 3583 CATGATACAGAAATTTGATGATGTCTGATCACAATAATGATTTTGTGATGAAAAAACT 3642
859 ----- 867
OY 3643 GATGTAGGATACGAAAAATATGCTTTAGAGATTAACACAGAGACATGATGTGCG 3702
868 ----- 868
OY 3703 TGTGGTGAAGTGTGATGCAAGACGACAGGTTACAGATTAACGAAAAAGAAAGCG 3762
869 CysGlyLeuLysTyrGluPro----- 875
OY 3763 GAAGAAAGATACGGAATGTAAACAGTGAATGATTAAGTAAAGAAAGAGAGAAAG 3822
876 ----- 876
OY 3823 AAACAAGTGAAGATTTGATCCAAAAAGAAATAGTATCCGATTTGCAATGC 3882

Db 879 GluLys----- 886
OY 3883 GGAATATATAATTTATG----- 3924
Db 887 ValThrProSerGlyValSerThrAlaThrSerGlyLysAlaValAlaIleCysValPro 906
OY 3925 CCTAGAAAGACAAAGTATTCGATCTTCTTGGCA----- 3969
907 ProArgArgArgLeuTyrValGlyGlyLysSerGlnTrpAlaSerArgLysLysP 926
OY 3970 GAAATAAAAAATTTACATACAGATTT----- 3999
927 GluThrThrGluValSerSerGlnAlaThrSerAlaProSerGlnSerGlnLys 946
OY 4000 TTTAAAGAACTTTCATCAATCTGCAGACAGAAACATCTCTCATGATTAAT 4059
947 LeuArgThrAlaPheIleGlnSerAlaAlaIleGlnThrPhePheLeuTrpHisLysTyr 966
OY 4060 AAA----- 4074
967 LysGluGluLysLysProAlaThrGlnAspGlyAlaGlyLeuGlyValSerLeuPro 986
OY 4075 ----- 4119
987 GluProSerProProGlyLysProGlnThrGlnLeuGlnThrGlyValIlePro 1006
OY 4120 CCCGATTTTTCAGATCCATGTTCTTACACATTTGAGATTAATGAGATTTTATTTGA 4179
1007 ProAspPheLeuArgLysMetPheTyrThrLeuAlaAspTyrLysAspIleLeu----- 1024
OY 4180 ACAGATATATCAAAAGTGTGAGGAGTAAATCAAAAGACCAATATGATTTCTT 4239
1024 ----- 1024
OY 4240 TTCAAAAATGTGACCAAAAATCTCTAATGAGAAAAACGCCAAGATGTGACAGAA 4299
1024 ----- 1024
OY 4300 CATAGTATGAGATTAATGGAAGCTATGCTATGTGACATGATTAATAATTTGGGCAAAAAA 4359
1024 ----- 1024
OY 4360 GATGATTTTACCGAAACTACGTTTACACACACGTCAAATTTAGTACAAAAAGACCACT 4419
1024 ----- 1024
OY 4420 TTGAGGAATTTGCCAAAGACCCAGTTTACGATGCTAACCGAATGTGACGACGAC 4479
1024 ----- 1024
OY 4480 TATGCTATACACGCAAAAATATTTGAGAGTGTGACGAAAAATGATGCAATGAC 4539
1024 ----- 1024
OY 4540 CAATTGAAGTGTGATACAGATTAATGAAGAAATGCGAGACTACGTTAATATATGAAA 4599
1024 ----- 1024
OY 4600 AAAAAAAGTGTGATTCACAAAGATTAATTTACAAGATGACCGCAAAAAAAGA 4659
1024 ----- 1024
OY 4660 TTCATAGACACACATTTGTATGTTATGATACAGACTATACGAAAGATGCAACAGAT 4719
1025 ----- 1032
OY 4720 TACTTGACAGAAATTTACTGCTAGTTGTGATGAAGCTGGAAGTGCCTCTGTGTA 4779
1032 ----- 1032
OY 4780 CAAGAAATATACAAATTTAGAAAAAAGAGCTTACTATGATGCCGCAAAATTTGTGG 4839

```
Db 1032 ----- 1032
QY 4840 TGCACAAATTTATGAAATGACGACAAATATACATACATTCGAGTAAGATGAC 4899
Db 1033 -----AspThrThrGlyLysGlnThrProSerSerAsnAspAsnLeu 1047
QY 4900 AAGAGTTACTAAGAGGCAACACAGGTCTATTAAAGTGCAAAACAAAGGCTCAAT 4959
Db 1048 LysAsnLeuValLeuGlnAlaSer ----- 1055
QY 4960 AACTACAATTAATCTGAAGAATTGACGAGATGTGCTTTCTCTCTGCGACTACGT 5019
Db 1055 ----- 1055
QY 5020 ATATGTTTCATCATGATTCGATGCAATTAATACATCCAGATCAAGATTAAGATGAATGG 5079
Db 1055 ----- 1055
QY 5080 TTCCGAAAAAGATTGATGAGTGGCGGCAAGGAGGATCAATTTGGTCAATACTAC 5139
Db 1056 -----GlySerThrGlu ----- 1059
QY 5140 AAGAAAAAAGAAAAAGAAAAATPAAAAACGTGCGATCGCAAAATATCTTANTAG 5199
Db 1060 -----GlnGluLysGluLysMetLys ----- 1066
QY 5200 GTCCCGCTGTAGTGTATGAATATATGTTTATGATTTAAGATATATCTAGGT 5259
Db 1066 ----- 1066
QY 5260 ATTGATTAATTTGGAGATGAAAAACAAAGACCGAGAAAAATTTGAAGAAATTTTAC 5319
Db 1067 -----GlnIleGlnAlaLysIleLysLysIleLeuAsn 1077
QY 5320 AAAAAATGACATCAGTTGGCAAGAGATAGTACTACAGAAATCCCGTAGACT 5379
Db 1078 -----GlyAlaThrSer -----GlyValProProValThrLysAsnSerValLysThr 1093
QY 5380 GCGCGAAATTTTCTCGAAGCAAAATTAAGCAATGTGTGCAACGCAATGATGCGGG 5439
Db 1094 ProGlnGlnThrTrpTrpGluAsnIleAlaLysAspIleTrpAsnAlaMetValCysAla 1113
QY 5440 -----TACCAACGTGTAGAGGATGATGAAATAGTGCAAAATAGTGCACAGTATGAA 5493
Db 1114 LeuThrTyrLysGlu -----AsnAspAlaArg ----- 1122
QY 5494 GATCTAAAAAATGTGTCTGTACCTTACATGATGATATATCTATGCGGAAAAATCGC 5553
Db 1122 ----- 1122
QY 5554 GATGAAGTACTGCTATATCAATTCCTTCGATGGTTGCGCAATGGGTGAAGATTTTTGC 5613
Db 1123 -----GlyThrSer ----- 1125
QY 5614 AAACATTAAGAAAAAGAAATTGAGAAATTTGTAAGGCGCTGTATATGATTACTTGTGT 5673
Db 1125 ----- 1125
QY 5674 GATTAATGAAGATAAAGAAAGAAATGTACAGATGCGTGTACACAAATTAATAATTTAT 5733
Db 1125 ----- 1125
QY 5734 AGTAGAGTAACACACAGATATGAAAAACAATCAAAAAATATGTGTGAATAAGACAAA 5793
Db 1126 -----AlaLysIleGlnGlnAsnLysAsp ----- 1133
QY 5794 ATATATTCGACGATCTCTGTGCAAAAGATGACAGAGACCGCTGCCAATATTTTAGACAAA 5853
Db 1133 ----- 1133
QY 5854 CAATTAATAAATAATTTGTGAATAATAAAGTGAAGATTGTGAATATATAGTATGAAGAT 5913
Db 1134 -----LeuLysLys ----- 1136

QY 5914 GTGTCCACACAGGATTAACGTAGTGAATAGTCAAAATATGCCCGCATATTAGACAT 5973
Db 1137 -----AlaLeuThrPaspGluAlaAsnLysAsnThrPro ----- 1147
QY 5974 GAACCAAAAGAAAGTTGAAGAAAGTGAATTTGTCAAGTGCACAGAGTCCACACAGTGA 6033
Db 1147 ----- 1147
QY 6034 CGAAGGGAACACCGTCCACACGCGTATCATGATATCAAAAGCAGCAGCATGCAAAAA 6093
Db 1147 ----- 1147
QY 6094 GAAGCGAAAAACAGCGCCCTACAAAAACCGCAAAAAAGTGGAAATCTAACACAGAA 6153
Db 1147 ----- 1147
QY 6154 ATGGAGCACAAACAGAACCCGACGACGACACAAACAGAAACAGAAACATCAACA 6213
Db 1147 ----- 1147
QY 6214 GCAACAAACACAGATCTGACGTGGGCAATGSTAAAGCCATTTCTTGAATTAACCA 6273
Db 1147 ----- 1147
QY 6274 GATAGCAGGGGTGAATAGAGGTTGTAAACAAACGTAAGACAAATATCCTAATGG 6333
Db 1147 ----- 1147
QY 6334 GGTGTATTGTAGTAAGTAAAGAAATGAAATGGCATATGATGCTCTCTAGAGA 6393
Db 1147 ----- 1147
QY 6394 AAAAAATATGTAATAATATATACATATTAATATGAACTGAATAAAGCTGAC 6453
Db 1148 -----IleGluLysTyrGlnTyrThrAsnValLysLeuGlu ----- 1159
QY 6454 AATGATTAATAAGAGCGCTTATTAATGTACAGCAATAGAACTCAATTTTGTGCTA 6513
Db 1159 ----- 1159
QY 6514 AATATATTAATGAAAAATCTGCAGCAGAAATGAATTCGAAATGGAACAATTCAGAT 6573
Db 1159 ----- 1159
QY 6574 GAATTAAGAAATATGATTAATACATATGATGATTAAGATATGTTTTTGAAGT 6633
Db 1159 ----- 1159
QY 6634 GATATTTCTAATGATAAAAAATTAATTAACGTGAACAAATAGTGTAACACATTCAT 6693
Db 1160 -----Asp ----- 1160
QY 6694 GAAATATTAAGAAAAACAGATAAAAAAGATGAAATTAAGTAAATATTTTGG 6753
Db 1161 GluSerGlyAlaLysSerAsnAsp ----- 1168
QY 6754 GAGAAAAATAAATTTATTTGGAGAGATGATATATGATTAATCAATTCATCACA 6813
Db 1168 ----- 1168
QY 6814 GAGGAAAAAGAAAAAGAAAAATTAAGATTAATACAGTACATGACATGACAAAACTG 6873
Db 1169 -----PheIleGln 1171
QY 6874 ACAGCTTCCCTGAAGAGTTTGAAGAAAGCCCAATTTTGAAGTGTTCACAGAAAG 6933
Db 1172 ProThrThrLeuLysAsnProValGlnIleProThrPhePheArgTyrLeuHisGluTrp 1191
QY 6934 GCAGAAATTTGTATTAAGAGCAAGACAGTGTGTAAATGTGAGGGGCGGTG 6990
Db 1192 GlyAsnSerPheCysPheGlnArgAlaLysArgLeuValIleGlnIleLysHisGluCysMet 1211
```


US-10-08/-013-1 (1-10628) X US-08-568-459A-10 (1-700)

Db 219 GluAsnLeuGlnLysGlnCysAspTyrGlnGlnAsnLeuTyrCysSerGlnLysnGly 233
 Oy 1069 CATGATGTTCAGACAACAATATTTGGAAAAAGCATATTTGGATTGGATTAATAGTACT 114
 Db 239 TTAAspCysThrLysThrLleTyrLysLysGlnLysLeuValIleGlyGlnHisCysThr 256
 Oy 1129 GACTGTGCACGTAAATGCAAAAGTTTGGAGTTGGTATGGAGATCAACAAAGACATTT 118
 Db 259 AsnCysSerValTTPCysAspMetCysGlnThrTTPrlAspAsnGlnLysLysGlnPhe 278
 Oy 1189 AAAAAACAAAAGAAAAATATGAAAAAGAAAT----- 1222
 Db 279 LeuLysGlnLysArgLysTyrGlnThrGlnLlleSerGlnGlyLysSerPro 298
 Oy 1222 -----CAATCATTTATTCGACAGATTAACAAATTTGTCAAT 125
 Db 299 LysAspThrLysAspGlnAlaIAspSerSerSerSerSerSerSerSerSerSerSer 314
 Oy 1258 AATATTAATAGTGAATATTTAAACAATTTTATGAAAACTTAAGAAACGCAATATGCA 131
 Db 315 -----GlyTyrGlnSerLysPheTyrLysLysLeuValGlyTyrGln 330
 Oy 1318 ACTATGACACTGTTTAAATTTACTAAATGAGAAAGATGTTGAAGA----- 1361
 Db 331 AspValAspLysPheLeuLysLlleLeuAsnLysGlnGlyLlleCysGlnLysGlnProGln 350
 Oy 1369 ---GGATTACACAGAGAAAGAAAGATATCTTTACTTACTACACTGCTGATGACAAAGGATA 1421
 Db 351 ValGlnLysAsnGlnLysLAspAsnValAspPheThrAsnGlnLysTyrValLys---Thr 369
 Oy 1426 TTTATATGTTCCAGAAATATTTGCCAAGTGTGTCCGACTCGGGGCTCAAAATGTATGTATA 1485
 Db 370 PheSerThrGlnThrLlleCysGlnProCysProThrPysGlnLeuGlnLysGlyGlnPro 389
 Oy 1486 AAAATCACACACAATCAATCAATATGATGATGCTGAACGTGTAATATGAAGCATTTAACT 1545
 Db 390 ProThrPheValLysGlnLysPheThrCysGlnGlySerAlaLysThrLysThrLysAspPro 409
 Oy 1546 CCATGGGCTGTGAAGCTCTTAATGTCACGTGCTCTTATATGCTGAATGACAAAGCTGAT 1605
 Db 410 Lys-----AsnIleThrAspLleThrValLleuTyrProAspLysSerGlnAsn 426
 Oy 1606 ATTACACAAAATTTAGAAAATTTTGTACACGCTCAACTAATTCAAAGATAAAATATAT 1665
 Db 427 ILeuLysLysTyrLysAsnPheCysGlnLysGlyAlaPro---GlyGlnGlnLle 445
 Oy 1666 CAAAAATGGGAATGCTATTAATAAGAT-----GAAATATTAATATGATGT 1710
 Db 446 LysLysTyrGlnCysTyrTyrAspGlnHisArgProSerSerLysAsnAsnAsnAsn 465
 Oy 1711 AAACGTGAAACAAATCTGCAAAATCAATATGATTAATCCTTAAGATATATCAATTTCAATAT 1770
 Db 466 ValGlnGlyThrTTPAspLysPheThrGlnGlnLysGlnThrValLysSerTyrAsnVal 485
 Oy 1771 TTTTGTGAATATGAGTAAATATATTAATGAAGATACATTAAGTGAATGACAAACTT 1830
 Db 486 PhePheThrAspThrValHisAspMetLeuHisAspSerValGlnTyrLysThrGlnLeu 505
 Oy 1831 AAAACTGTATAATATATACACCAG-----CATGTATTAATGAGATTAAC 1878
 Db 506 SerLysCysLleAsnAsnAsnThrAsnGlnLysAsnThrCysArgAsnAsnAsnLysCysLys 525
 Oy 1879 AGAAATTCCTTAATGTTTGCAGATGGGTTTAAACAAAAGAAAGAAAGATGGAATAGTATA 1938
 Db 536 ThrAspCysGlnLysPheGlnLysThrValGlnLysLysGlnGlnGlnLysPheMetAlaIle 545
 Oy 1939 AAGAACTGTCAAAAAGAAATATACAGCAATCGATATATAGTATATTAATAT 1998
 Db 546 LysAspHisThrLlleLysGlnHisAspLleValGlnGln-----LysGly 560
 Oy 1999 CTTTTCGAACGTTATTTTATTAATGTTATGATTAACCTGACAAAGATCAACAAATGCG 2058
 Db 561 LeuIleValPheSerProThrValValThrThrThrThrThrThrThrThrThrThrThr


```

QY 2059 AAGAACTTTCGAAATATATAAAGAAAAATGAGTTTCCATTTCGAAATAT 2118
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 577 GYASLLEULEUNGINSNILEYASPVAlHISGLYASPhrAspRIleYSHISile 596
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 2119 AGGACATTTAGAG-----AT 2136
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 597 LYSLEULEUNSPGLUGLIALAVALAlAVALleUGLYLYASpASpRnThr 616
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 2137 GCATAGAACTCTGTAGACTTAAAGAACTCCAGATGTAAGCAATAT 2196
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 617 ThrILeAspLYLeULEUNGINSGLULYGLUGLIALeUGLYLYSGLNLYS--- 635
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 2197 ACAAGACAGCATGTCAA-----ACATCCCAATTAATGCAACACA 2235
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 636 ---GlnGluGluCysGluLYSLYSAlAGlnGluGluSerArgLYASpSerAlAGluThr 654
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 2236 AACCCGCTGTAAACCTGTCGAGCAGCAGCCCACTAAATATAAAGAAATGCA 2295
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 655 ARGGLuAspGLUArgTYrGln-----GlnProAlAspSerAlAGLYLValGlu 671
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 2296 CAATACTTTAAAGAGTGCATACGAGAA 2325
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 672 GlnGluGluAspAspAspTYrAspGlu 681
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 7
US-08-487-826B-10
; Sequence 10, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Slim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-487-826B-10

```

```

Alignment Scores:
Pred. No.: 2,36e-70 Length: 700
Score: 1032.50 Matches: 247
Percent Similarity: 50.28 Conservative: 110
Best Local Similarity: 34.79 Mismatches: 250
Query Match: 5.318 Indels: 103
DB: 2 Gaps: 22

US-10-087-013-1 (1-10628) x US-08-487-826B-10 (1-700)
QY 406 AATGATGCTTACCTCTGCGCCACCAAGACGACATGTGTGTAATAACTGGAA 465
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 5 AsnLYSValGLYAlAcYSAlAPrOTYrArgLYLeuHISLeuYSAspTYrAsnLeuGlu 24
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 466 GCTCTAAATGATTAATATCCCAAAATATCATGATTATTTGGAAATGACTAGTACA 525
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 25 SerILeAsp-----ThrThrSerThrHisLYSLeuLeuGluValLYSmetAla 42
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 526 GCAAAATACGAGAGTGAAATCAATGTTTAATATCAT-----CCACATAAGGAACT-- 576
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 43 AlAlSTYrGLUGLYAsnSerILE---AsnThrHisTYrThrGlnHISGlnArgThrAsn 61
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 577 -----TCAGACGCTTGACTGCTTGTCGACGAAGTTTGGACATATAGTGAT 624
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 GluAspSerAlAspGlnLeuCysThrValLeuAlAspSerPheAlAspRIleGlyAsp 81
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 625 ATGTGAGAGCAATGATATGTT-----AAACCAATGTCCATGAC 666
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 82 ILleValrGLYLYSAspLeuTYrLeuGLYTYrAspAsnLYSGLULYsGLULNArgLYs 101
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 667 AAAGTAGAAACGGGTCGCCAGAGGTTTCAAGAAATATACATGATGAAATGAGAA 726
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 102 LYSLeuGluGlnLYSLeuLYSAspRIlePheLYSLeuHIS-----LYSAspVal 118
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 727 GTAAAAATGAT-----TACAATCCATGACA---TCTGCAATTTATAT 768
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 119 MetLYSThrAsnGLYAlAGlnGluArgTYrILeAspAspAlAspGLYGLYAspPhe 138
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 769 AAATTAAGAGACATGTCGATGTAATGAAATTAATGGAAGCTATATACATGT 828
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 139 GlnLeuArgLYSAspTrpThrPheAsnArgLYSValThrValTrpLYSAlAlleCYs 158
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 829 GATGCATCATATTAATCTGATATTTTATGCAATACAGAAAGTATACCATATTTTCA 888
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 159 HISAlAPrOLYsGLUAlAsnTYrPheILeLYSThrAlAcYSAsnValGLYsGLYThr 178
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 889 AATCTTAATGCCGCCATAACAAGCAAGATTCCTACCAATTAGATTATGCCCTCAA 948
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 179 AsnGLYGLNcYSAspIleGLYGLYAspValPrOThrTYrPheAspTYrValPrOTGln 198
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 949 TATTACGTTGGTTCGAGCAATGGGAGAGAGTTTGGCGAAAGAAATATATTAATG 1008
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 199 TYrLeuArgTrpPheGLUGluGluTrpAlAGluAspPheCysArgLYSLYSLeu 218
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1009 AAAAAGTCAAGACATCTCTCGTAATGACAAGACGTTAATGCTACTCATATGCA 1068
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 219 GluAsnLeuGlnLYSGLNcYSArgAspTYrGLUGlnAsnLeuTYrCysSerGLYAsnGLY 238
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1069 CATGATGTACGACAACTATTGGAAAAAGCATTTTGCATTTTGCATTAATGACT 1128
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 239 TYrAspCysThrLYSThrILeTYrLYSGLYsLYSLeuValILleGLYGLNHISCYsThr 258
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1129 GACTGTTGACTAAATGCAAGATTTTGAAGTTTGAAGATGAGCAACAAGAGCTT 1188
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 259 AsnCYsSerValTrpCYsArgMetTYrGLUThrTrpILeAspAsnGlnLYSGLU 278
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1189 AAAAACAAGAAAGAAATATGAAAAAGAAATA-----1221
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 279 LeuLYSGLNLYSArgLYSTYrGLUThrGluILeSerGLYGLYsGLYsSerPro 298
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1222 -----CATCATATTATCGAACCATATATTTGCAAT 1257
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```

Db      25 SerIleasp-----ThrtTserThrThrHisLysLeuLeuGluValCysMetAla 42
OY      526 GCAAAATACGAGGTGAATCAATGTTAATATCAT-----CCACATAAGCAACT--- 576
Db      43 AlalysIyrgIuGluLysnsrIle--AsnThHisIyrgThrgInHisIahIyqThAsn 61
OY      577 -----TCAGAGCTGTACTGCTCTTCGACGAAAGTTTGGATATAGTATG 624
Db      62 GluAspSerAlaSerGlnLeuCysThrValLeuAlaIargSerPheAlaAspIleGlyAsp 81
OY      625 ATTGTAGAGGAATAGATATGTTT-----AACCAAAATGCCATGAC 666
Db      82 IleValAlrgIuLysAspLeuIyrgLysIyrgLysAspAsnLysGluLysGluIahIyrgLys 101
OY      667 AAAGTGAAGAACGGGTCTCCGAGAGGTTTTCAGAGAAATATACATGATGAAATGAGCAAGAA 726
Db      102 LysLeuGluGluLysLeuLysAspIlePheLysIleHis-----LysAspVal 118
OY      727 GTAAAAAATGAT-----TACAAATCTGATGGA---TCTGAAATTTATAT 768
Db      119 MetLysThrasnGluIahIagInIuIargIyIleAspAspAlaLysGluLysAspPhePhe 138
OY      769 AAATTAGACAGACATGCTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 828
Db      139 GlnLeuAlrgIuLysAspIlePheIyrgLysAspIlePheIyrgLysAspIlePheIyrgLys 158
OY      829 GATGCATCATATATGATATGATATGATATGATATGATATGATATGATATGATATGATATGAT 888
Db      159 HisAlaIProLysGluLysAlaAsnIyrgLysIleLysThrAlaCysAsnValGlyLysGlyThr 178
OY      889 AATCTTAATATGCGCCATTAACAGAGAAAGTTCTTACCAATTTAGATTTAGTCCCTCA 948
Db      179 AsnGluLysGlnIyrgLysIleGlyLysAspValProIyrgLysPheAspIyrgLysProGln 198
OY      949 TATTTACGTGTGTTCAGACAGATGGGAGAGAGATTTTGGCGAAAGAAATATTAATG 1008
Db      199 TyrLeuAlrgIyrgLysGlnIyrgLysAlaAspPheCysAspIyrgLysLysLysLeu 218
OY      1009 AAAAAGTCAAGAGACATCTGTAATGACAAAGACGTTTATGTTATGTTATGTTATGTTATGTT 1068
Db      219 GluAsnLeuGlnIyrgLysGlnIyrgLysAspIyrgLysGlnIyrgLysGlnIyrgLysGln 238
OY      1069 CATGATGTGACGACATATTTGGAAGAAAGATTTTGCATTTGATTAATAGTGTACT 1128
Db      239 TyrAspCysThrLysThrIleTyrLysGlyLysLeuValIleGlyLysIahIyrgLys 258
OY      1129 GACTGTTCCGACTAAATGCAAGTTTGAAGTTTGTGAGGAATCAACAAGAACGTTT 1188
Db      259 AsnCysSerValIyrgLysAspGlyGluThrIleAspAsnGlnLysGluPhe 278
OY      1189 AAAAACAAGAAAGAAATATGAAAGAAATA----- 1221
Db      279 LeuLysGlnIyrgLysIyrgLysIyrgLysIyrgLysIyrgLysIyrgLysIyrgLys 298
OY      1222 -----CAATCATATTTATGACAGCAATCAAAATTTGTCAT 1257
Db      299 LysArgThrLysAlaIahIahIahIahIahIahIahIahIahIahIahIahIahIahIah 314
OY      1258 AATATTTAATGTGATATTTAACAATTTTGAAGAAAGTTAAGAAAGCAATATGCA 1317
Db      315 -----GlyTyrGluSerLysPheTyrLysLysLysGluValGlyTyrGln 330
OY      1318 ACTAATGACACATTTTAAATTTACTAATGAGAGAAAGATTTGTAAGA----- 1368
Db      331 AspValAspLysPheLeuLysIleLeuAsnLysGlnIyrgLysGlnIyrgLysGlnIyrgLys 350
OY      1369 ---GGATTACAGAGAGAAAGATATTACTTTACTAACAAGTGTGATGCAAGAGGATA 1425
Db      351 ValGlyAsnGlnIyrgLysAlaAspValAspPheThrAsnGlnIyrgLysIyrgLys---Thr 369
OY      1426 TTTATCGTTCAGAAATTTCCCAAGTGTCCGACGCGGGGCAAAATGATGATGATATA 1485

```

```

Db      370 PheSerArgThrGluIleCysGluProCysProIyrgLysGluLysGlyPro 389
OY      1486 AAATPACACACAAATCGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1545
Db      390 ProIyrgLysValLysGlyAspLysThrCysGlySerAlaLysThrLysThrLysAspPro 409
OY      1546 CCATGGGGGTGGAAGCCCTAATATATCAGTCTGTTTATGATGATGATGATGATGATGATGAT 1605
Db      410 Lys-----AsnIleThrAspIleProValLeuIyrgLysAspLysSerGlnIahIyrg 426
OY      1606 ATTACACAAAATTTAGAAATTTTGTACAGCTCAACTAATTTACAAAGATTAATAATTAAT 1665
Db      427 IleLysLysIyrgLysAspPheCysGluLysGlyAlaPro---GlyGlyGlyGlnIle 445
OY      1666 CAATAATGGAAGATCTTATTAAGAT-----GAAATATTAATTAAGATGT 1710
Db      446 LysLysIyrgLysIyrgLysIyrgLysIyrgLysIyrgLysIyrgLysIyrgLysIyrgLys 465
OY      1711 AAATGGAACAAAATATGATCAATATGATATGATATGATATGATATGATATGATATGATAT 1770
Db      466 ValGluGlyThrIyrgLysLysPheThrGlnGlyLysGlnIyrgLysIyrgLysIyrgLys 485
OY      1771 TTTTGTGAATATGCGTTCATATATTTATTAAGGATCTTAAAGTGAATGACAAACT 1830
Db      486 PheThrAspIyrgLysIyrgLysIyrgLysIyrgLysIyrgLysIyrgLysIyrgLysIyrg 505
OY      1831 AAACCTGTATTAATATATACACACAG-----CATGTATGATGATGATGATGATGAT 1878
Db      506 SerLysCysIleAsnAsnAsnIyrgLysIyrgLysIyrgLysIyrgLysIyrgLysIyrgLys 525
OY      1879 AGAAATGCTTATGTTTTCAGACATGCTTAAACAAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1938
Db      526 ThrAspCysGlyCysPheGlnIyrgLysIyrgLysIyrgLysIyrgLysIyrgLysIyrgLys 545
OY      1939 AAGAACTGTTCACAAAAGAAAGAAAGAAATATACAGCAATGATATTAATTAATTAAT 1998
Db      546 LysAspIlePheGlyLysGlnIyrgLysIyrgLysIyrgLysIyrgLysIyrgLysIyrgLys 560
OY      1999 CTTTTGAAGCTTATTTTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2058
Db      561 LeuIleValIahPheSerProIyrgLysValLeuAspLeuVal-----LeuLysGly 576
OY      2059 AAAGAACTTATGCAAAATATTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2118
Db      577 GlyAsnLeuGlnIahIahIahIahIahIahIahIahIahIahIahIahIahIahIahIah 596
OY      2119 AGGAGCTATTTAG----- 2136
Db      597 LysLysLeuLeuAspGluGluAspAlaValAlaValIahIahIahIahIahIahIahIah 616
OY      2137 GCAATGACACTCTGTTAGATCACTTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2196
Db      617 ThrIleAspLysLeuLeuGlnIahIahIahIahIahIahIahIahIahIahIahIahIah 635
OY      2197 ACAAAAGACATGTGA-----ACATCCCAATATGCAACACA 2235
Db      636 ---GlnGluLysGlyLysIyrgLysAlaGlnIyrgLysIyrgLysIyrgLysIyrgLysIyrg 654
OY      2236 AACCCGTGTGTAAACCTGTCGAGGACCAACCAACCAACCAACCAACCAACCAACCAACCA 2295
Db      655 ArgGluAspGluIyrgLysIyrgLysIyrgLysIyrgLysIyrgLysIyrgLysIyrgLys 671
OY      2296 CAATCATTTAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2325
Db      672 GluGlnIyrgLysAspAspIyrgLysIyrgLysIyrgLysIyrgLysIyrgLysIyrgLys 681

```

RESULT 9
 US-08-568-459A-8
 ; Sequence 8, Application US/08568459A
 ; Patent No. 5849306
 ; GENERAL INFORMATION:
 APPLICANT: Sim, Kim L.
 APPLICANT: Chltnis, Chetan

```

APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhan
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE: Plasmodium falciparum
US-08-568-459A-8

Alignment Scores:
Pred. No.: 3,63e-41
Score: 649.50
Percent Similarity: 39.48%
Best Local Similarity: 26.18%
Query Match: 3.34%
                2
                Gaps: 48

-10-087-013-1 (1-10628) x US-08-568-459A-8 (1-921)
QY 2842 GATATATTCGAGAGAGATCTCTGGAGAAAGCGTGACATGCTGAAGCTGCAAGA 2901
DB 80 Asptylrllthralsersaplle--gluasnglyglasnserrlle----- 94
QY 2902 CATTGGAAACGTTTGGTAATATACCTAAAGTCACTCAAGGCAAGAAATGATAA 2961
DB 95 -----Glyasnileaspmetvalvalserasplyasplasnngly 108
QY 2962 TATTAAT-----GATGATGCCGCCCAATATTTAAATTTGAGGAAATTTGG 3006
DB 109 Pheasnnglyleuasplalcysclysersalasnilepnelysglyllearglyslngln 128
QY 3007 TGGGAAGCTAATAGAGCCAAAGTATGGAAAGCATGAATGTGATATAATATTGAG 3066
DB 129 Trplyscys-----Alalysval-----Cysglyleuasplalcyscl 141
QY 3067 GATATTCGGGACACCAATCAACACAAAGTATTGCGGATATAGTATACATACACA 3126
DB 142 Leuylsasnglyasn-----Glyserilleasplyaspln 153
QY 3127 TTGATGATATATCCACAAATAATAGATGATGACCAAGTGGCAAGTGTACTGC 3186

```

```

DB 154 Lysglnlellellelleargalaleu-----Leuylsargtrpvalglutyrphee 171
QY 3187 AAGTGCAGAAAAGAGTATGATTAAGTGAAGAGAGTGAAGCTGAAGTAAAG 3246
DB 172 -----Gluaspyrlyasnllysleasnalellysleaserlscysrhllys 187
QY 3247 GATATTCGCAAGGCTGTGAGAAAGAGCTGACAGGTTGTACGAAGTGCACAGAGCT 3306
DB 188 Aspasnnglyserthrscysrhlrasnaspcysproasnlyscysrhlr---Cysvalglu 205
QY 3307 TGTATGATATATATGATATATATGATATATGATATATGATATATGATATATGAT 3366
DB 206 -----GluTrpilleasnnglnlyasargthrglutrpysasnilleyslys 220
QY 3367 AATATACAAAGATTCATGAA----- 3387
DB 221 HistyllystrhnglnasnnglnnglnlyasnasnasmetlyserleuValthrspile 240
QY 3388 -----CAAGCAAAATGCTGTCTGTTACTAAT-----AGTGTATTT 3420
DB 241 LeuglylaleuInProdlInseraspyalasnlysalalleysProcyserglyleu 260
QY 3421 GAAGCT-----TTCAGTACGCCAAAATCAT 3447
DB 261 ThrAlaPhegluserPheCysglyleuasnglyalalaspasnsersclulysglngly 280
QY 3448 ATAGAC-----AGAAATGTTATGTAATTTTGTGCAATTA 3483
DB 281 GluaspyrlyaspleuValleucysmetleuysnasleuInglulysglnleaglnclys 300
QY 3484 TACCAACAATAATGTGCGCAAAAGTAAATGAAGTGTACTAGTGAAGAGCTGTCAAT 3543
DB 301 LyslyslslyshlsclglutrhserValalulasnngly-----GlylysSerCysThrPro 318
QY 3544 GGTACTAACCCAGTATGAAATGTTGACACATATCTCATGATACAGAAATTTGAT 3603
DB 319 LeuaspaasnThrThrleuglglu-----Glu 327
QY 3604 GATGTCTAGTCAAAAATGAG-----TTTGTGATGAAAAAGT----- 3642
DB 328 ProileglnlgluIngluasnglnvalgluAlaProasnllcysProlysglnThrValglu 347
QY 3643 GATGTGAAGTAAAGCAAAA-----TATGCTTTAGAGNTAAA 3681
DB 348 AsplyslslysglnlgluInglulthrcysThrProalaserProvalProglulys 367
QY 3682 CCACAGACCATGATGCTGCTGTGTTGAAAGTGAAGTGAACCGACAGAGTACAG 3741
DB 368 ProvalProhis-----ValAlaArgTrpArgThrPheThrProProgluValPheLys 385
QY 3742 ATAAAAAGAAAAAAGCGAAGAAAGATAGCAAGTGAACAGATGATATTA 3801
DB 386 Ile-----TrpArggluArgArgaslyslThrCysgluileValAlaIglumet 402
QY 3802 CTTAAGAAAGATGGAAGAAAGAAAGTGAAGATGTGATCAAAAAGATTAAT 3861
DB 403 LeuylsasplysasnglyargthrthValgllycysgluileValAlaIglumet 420
QY 3862 GATATTCGATTCGCAATGCGCAAAATATTAATTTA-----GTGGAAGACCTGTGTGT 3918
DB 421 ---Tyserslutrphrthrcysaspluserllyleymetclylnglnsllyalacs 439
QY 3919 ATGCCCCCTTGAAGACAAAGTATGCTGATTTCTTGGCAAAAGATATGAATAA 3978
DB 440 IleProProArgArgglnglnlyleucysleuHlstryleu-----Glu 453
QY 3979 AATTTCAATCAAAATTAAT---TTAAAGAAGCTTTCATCAATTCGACAGACAGAA 4035
DB 454 LysllemetrhthasnThrasnThrasnThrasnThrasnThrasnThrasnThrasn 473
QY 4036 ACATTCCTCTCATGTAT---TATTATTAAGTAAAGTGAAGTGAAGAAAGTGCAT 4092

```

```

Db 474 ThrPheLeuLeuTrrpGlnAsnTrrLysAspLysAsnGlyAsnAlaGluAspLeuAsp 493
QY 4093 AAGAAATTAAGAGAGCAAAATTCCTCCGATTTTGGATCCATGTCATACATTT 4152
    ::::::::::::::::::::
Db 494 GluLysLeuLeuGlyGlyLeuLeuProGluAspPheLysArgGlnMetHeterThrPhe 513
QY 4153 GCGAATTATAGAGATTTTATTGGAACAGATATA-----TCAAAA 4194
    ::::::::::::::::::::
Db 514 AlaAspTrrArgAspLleCysLeuGlyThrAspLleSerSerLysLysAspThrSerLys 533
QY 4195 GGTATGGTGGAGGAGAACTAAAGCAAAATGATTCCTTTTCAAAAATGGTGCAC 4254
    ::::::::::::::::::::
Db 534 GlyValGly-----LysValLysCysAsnLleAspAspValPhe----- 546
QY 4255 CAAAAATCTCTAATGA---AAACACGCCAAGAAATGGTGCACAGACATGTCATGAG 4311
    ::::::::::::::::::::
Db 547 TrrLysLleSerAsnSerLleArgTrrArgLysSerTrrPrrGluThrAsnGlyProVal 566
QY 4312 ATATGGGAAGCTATGCTATGTCACATA-----GTAAAAAT 4347
    ::::::::::::::::::::
    567 lIeTrrPrrGluGlyMetLeuCysAlaLeuSerTrrAspThrSerLeuAsnAsnValAsnPro 586
QY 4348 GGGGCAAAAAAAGATGATTTTACCGAAAACCTACGGTTACACAGCTCAAAATTT---AGT 4404
    ::::::::::::::::::::
Db 587 GluThrHisLysLysLeuThrGluGlyAsnAsnAsnPheGluLysValLlePheGlySer 606
QY 4405 GACAAAAGCACACTTGGAGGAAATTTGCCAAACGACCCAGTTTTCAGATGGCTACAC 4464
    ::::::::::::::::::::
Db 607 AspSerSerThrLleuSerLysPheSerGluArgProGluPheLeuArgTrrPrrLeuThr 626
QY 4465 GAATGGTACGACGACTATTCCTATTCATACGACACAAAAATATTGAGAGATGTCACGAAAAA 4524
    ::::::::::::::::::::
Db 627 GluTrrPrrGluLysAsnLysGluGlyLysGluTrrLysValLeuLeuAlaLys 646
QY 4525 TGTATGTCAAATGACCAATTTGAAGTGTATACAGATGTAATAGAA-----TGC 4575
    ::::::::::::::::::::
Db 647 CysLysAspCysAsp---ValAspGluAspGlyLysCysAsnGlyLysCysValAlaCys 665
QY 4576 GAGGACTACGTTAA-----TATATGAAA 4599
    ::::::::::::::::::::
Db 666 LysAspGluCysLysGlnTrrHisSerTrrPrrGlyLleTrrPrrLeuAspAsnTrrLysLys 685
QY 4600 AAAAAAAGATGCGATTCGCACACAGATAA-----TATTACAGAGATGACCGCGCAAA 4653
    ::::::::::::::::::::
Db 686 GlnLysGlyArgTrrGluValLysLysLleProLeuTrrLysGluAspLysAspVal 705
QY 4654 AAAAGATTCGAT-----AGACACACATTTGGTGTATGCTTACAGACTATACGAGACG 4707
    ::::::::::::::::::::
    706 LysAsnSerAspAspAlaArgAspTrrLysLysThrGlnLeuGlnAsnMetLysCysVal 725
QY 4708 AATGCACAGATTTACTTGAAC-----AGCAAAATTTTACTGCTACTTGT 4749
    ::::::::::::::::::::
Db 726 AsnGlyThrThrAspGluAsnGlyLysGluTrrLysCysMetHisLysPrrSerSerThrAsn 745
QY 4750 GGTGATACCCCTGGAGAGCTCTGTGTGTACAAAGAAATATACAAATTTGTTAAAAACAG 4809
    ::::::::::::::::::::
Db 746 SerAspMetProGluSer-----LysAspGluLys 755
QY 4810 GCTTACTATGATGCGCAAAACATTTGGTGCACAAAAATTTATGAAATGACGACAA 4869
    ::::::::::::::::::::
Db 756 ProGluLysValLysAspLys---CysAsnGlyCys---ValProAsnGluCysAsn 771
QY 4870 TATACTACATTTTCGAGT----- 4887
    ::::::::::::::::::::
Db 772 AlaLeuSerValSerLysSerGlyPheProAspGlyGlnAlaPheGlyGlyValLeu 791
QY 4888 AAGATTAAGTGCAGAAAGATTTAGTAAGGAGGCAACAGGTCTTAAAGTGGCAAAAC 4947
    ::::::::::::::::::::
Db 792 GluGlyThrCysLysGlyLeu-----GlyLysProLys---LysLys 804
QY 4948 AAGGTCTCTAATACATTAACATTTGAAAGAAATGACTGAG----- 4990
    ::::::::::::::::::::
Db 805 lIeGluProGluTrrLysPrrThrAsnAspLle-LeuLysSerThrLleProValTh 824

```

```

QY 4990 ----- 4990
Db 824 lIeValLeuAlaLeuGlySerLleAlaPheLeuPheMetLysValLleTrrLleTrrVal 844
QY 4990 ----- 4990
Db 844 lTrrPrrLleTrrMetLeuCysValGlyAlaLeuAspPrrTrrLleCysGlyCysLleCys 864
QY 4991 -ATGTCCTTTTCCTCTCGTCGAC-----TAGCAATATGTTTCATGCA 5034
    ::::::::::::::::::::
Db 864 sIleCysLlePheLleCysValSerValTrrValCysValTrrValTrrValPheLeuTrr 884
QY 5035 T-----TGATGGCAATTAACAGATCCAGATCAAGATTAAGATGAAAT 5076
    ::::::::::::::::::::
Db 884 rMetCysValPheTrrLleTrrPheLleTrrLleTrrLleTrrLleLeuLysMetLysLys 904
QY 5077 GGGTTCGAAAAAAGATTGATGAGAGTGGCGCAACGGAAGGTTACATTTGGTCAATAC 5136
    ::::::::::::::::::::
Db 904 s-----MetLysLys 907
QY 5137 TCAAAAGAAAAAAGAAAAAGAAAAATAAAGCGTGGAG 5179
    ::::::::::::::::::::
Db 907 sMetLysLysMetLysLysMetLysLysArgLysLysArgLys 921

RESULT 10
US-08-487-826B-8
; Sequence 8, Application US/08487826B
; Patent: No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-487-826B-8

```

Alignment Scores:

Pred. No.:	3.63e-41	length:	922
Score:	649.50	Matches:	250
Percent Similarity:	39.488	Conservative:	127
Best Local Similarity:	26.188	Mismatches:	299
Query Match:	3.348	Indels:	48
DB:	2	Gaps:	28

US-10-08/-013-1 (1-10628) x US-08-487-826B-8 (1-921)

[illegible]

D	b	328	ProIleeluglucLusnblvalGelualaProlasnlleCysPrdlysglnthrValglu	347
OY	3643	GATGCTAGGATACAGAATAA-----TATGGCTTTAGACATATAA	368:	
D	b	348	AspLyssylLysglucLueLuglunGluthrCysThrProLasSerProvalProgluLys	367
OY	3682	CCACAGACCACATGATGTGCCTCGTGTTGAAGTGGATCGAAACCGACAMGGTCACG	374:	
D	b	368	ProvalProHis-----ValAlaArGrTrpArgThrRheThrProProgluValPhelys	385
OY	3742	ATTAATAACAAAAAAGAACCGAGAAAAGCATCGCATGTAATAAACGATGATGATA	380:	
D	b	386	Ile-----TrpArgglYargArgLysnlysrThrCysgluleValalaeLuwet	402
OY	3802	CTTAAGCAAAAGATGATGAAAGAAACAAGTAGACAGATTGTGATCCAAAAGAAATGTAT	386:	
D	b	403	LeuLysAspLysasnGlYargThrThrValGlYelucLysTyArgLysgluthr-----	420
OY	3862	GGATATCCGATTTGGCAATGGGAAATATATAATTA-----GTGGAAAGACCCTGGTGNGT	3918:	
D	b	421	---TyrSerGlutrrThrCysAspGlserLysLileLysmetClYelnHIselYalaCys	439
OY	3919	ATGCCCCCTAGAAACAAATAATTTATGGCTGCATTTCTTGCCAATGATGATGAATAATAA	3978:	
D	b	440	ILleProProHArgArgGlInLYlsLeyCysLeuHIsTyLeu-----Glu	453
OY	3979	AAATTACATCAACAAGTTAAT---TTAAAGAAAGCTTATCATCAATTCGACACACACAA	4035:	
D	b	454	LysIleMetThrAsnThrAsnGlLueLysTyArgLisRheIleLysCysAlaAlaIeLu	473
OY	4036	ACATCTCTCTCATGGTAT---TATTATAAAGTAGAGATGTGAAGAAATGCACTCGAT	4092:	
D	b	474	ThrPheLueLueurTrpGlnAsnTyLysLysAspLysasnGlYasnAlaGlUnaspLeusp	493
OY	4093	AAAGATTTAAAGAGACCAAATTCCTCCCGCATTTTGTGATTCATCTGTCCATGACTT	4152:	
D	b	494	GlutylsLeuLysglYelYlleIleProglUnasRheLysArgLimeCheryThrPhe	513
OY	4153	GGAGTTTATAGACGTTTTTTTATTTGGAACAGATATA-----TCAAAA	4194:	
D	b	514	AlaAspTyArgAspRIleCysLeuGlYThrAspIIseSerLysLysAspThrSerLys	533
OY	4195	GGTCATGGTGGAGGAAATTAACATACTAAAGACAAATAGATTCCTTTCAAAAATGGTGAC	4254:	
D	b	534	GlyValaGly-----LysValLysCysasnIleaspAspValPhe-----	546
OY	4255	CAAAATCTCCTTAATGGA---AAACACGCCCAAGAGTGGTGACAGACATAGTCATGAC	4311:	
D	b	547	TyrlYslIleSerasnSerIleargTyArgLysSerTrpTrpGlunThrAsnGlProval	566
OY	4312	ATATGGAGACCTATGGCATGTGCACAA-----GTAAAAATT	4347:	
D	b	567	ILletrpLunclYmetLueCysAlaLueSerTYraspThrSerLueAsnAsnValasnPro	586
OY	4348	GGGGCAAAAAAGATGATTTTACCGCAAACTACGGTTACCAACAGTCAAAATTT---AGT	4404:	
D	b	587	GlunThrHisLysLysLueThrGlnGlYasnAsnAsnPhelGluLysValIlePhelGlySer	606
OY	4405	GACAAAAAGCACCTTTGGAGAAATTTGGCAAAAGCAGCCCCAGTTTTCATGATGGCTAAC	4464:	
D	b	607	AspSerSerThrThrLueSerLysPheSerGlnArgProglInPheLueAgtTrpleuthr	626
OY	4465	GAATGCTAGACAGACTATTTGCTATACAGACAAAAAATTTGTAAGAGATGTGGAGGAAAA	4524:	
D	b	627	GlunTrpTrpYglUnasphCysLysglunGlunLysLysglunTyrlYsnValLueLueAlaLys	646
OY	4525	TGTATGTCAAATGACCAATGTAAGTGTGATACAGATATGTATATAAGAA-----TGC	4575:	
D	b	647	CysLysAspCysasp---ValaspelYaspelYlYscYasnGlYscYsValaLaCys	665
OY	4576	GAGACTACAGTTAAA-----TATATGAAA	4599:	
D	b	666	LYsaAspGlCysLysglunTyrlHisSerTrpIleelYlletrpLleaspsanTyrlYsls	685

Db 188 AspAsnGluSerThrCysThrAsnAspCysProAsnLysCysThr---CysValGlu--- 205
 QY 3307 TGTATGAAATTAATGATTAATGATTAATGAAAGCAATGAAATTAATATGATGAT 3366
 Db 206 -----GluTrpIleAsnGlnLysArgThrGluTrpLysAsnIleLysLys 220
 QY 3367 AAATACAAAGAAATTACATGAA----- 3387
 Db 221 HisTrpLysThrGlnAsnGlnAsnGlnLysAspAsnAsnMetLysSerIleValThrAspIle 240
 QY 3388 -----CAAGACCAAAATGTCTGTAGTAAT-----AGTGTAT 3420
 Db 241 LeuGlyAlaLeuGlnProGlnIleSerAspValAsnLysAlaIleLysProCysSerGlyLeu 260
 QY 3421 GAAGCT-----TCCAGTACTGCCCAAAATATCAT 3447
 Db 261 ThrAlaIleGluSerPheCysGlyLeuAsnGlyAlaAspAsnSerGluLysLysGly 280
 QY 3448 ATAGAC-----AGCAATGTATTGAATTTTGTCCGAATTA 3483
 Db 281 GluAspTrpAspLeuValLeuCysMetLeuLysAsnLeuGlnLysGlnIleGlnLys 300
 QY 3484 TACCAAAATAATGTGGCAAAAGTAATAAGTGTACTAGTAAAGTGCCTGAT 3543
 Db 301 LysLysLysHisGlyGluThrSerValGluAsnGly-----GlyLysSerCysThrPro 318
 QY 3544 GGTACTAACACACCGTAAATGTTGACACATATCTCATGTATACAGAAATTTGAT 3603
 Db 319 LeuAspAsnThrThrLeuGlnGlu-----Glu 327
 QY 3604 GATTGTCACTCAAAATAGAC-----TTTGTGATGAAAAAGT----- 3642
 Db 328 ProIleGlnGlnGlnGlnGlnValGlnAlaProAsnIleCysProLysGlnThrValGln 347
 QY 3643 GATGTTAGATTAACGAAAA-----TATGCTTTAGAGATTA 3681
 Db 348 AspLysLysLysGlnGlnGlnGlnGlnGlnCysThrProLaseProValProGlnLys 367
 QY 3682 CCACAGACATGATGTGCTGTGTTGTAAGAGTGAACCGAACCGACAGGTACAG 3741
 Db 368 ProValProHis-----ValAlaArgTrpArgThrPheThrProGluValIlePheLys 385
 QY 3742 ATAAAGAGAAAAAGCGGAAAGAAAGCAAGATACGAAATGTAACAGTATGATTA 3801
 Db 386 Ile-----TrpArgGlyArgArgAsnLysThrThrCysGlnIleValAlaGluMet 402
 QY 3802 CTAAAGAAAAAGTGAAGAAAGAAAGTGAAGATGTGCATCCAAAAAGATAGTAAT 3861
 Db 403 LeuLysAspLysAsnGlyArgThrThrValGlyLysCysLysArgLysGluThr----- 420
 QY 3862 GGATATCCCATGGCAATGCGGAAATATTAATTA-----GTGGAAGACCTCGTGTGT 3918
 Db 421 ---TysSerGlnTrpThrCysAspGluSerLysIleLysMetGlnHisGlyAlaCys 439
 QY 3919 ATGCCCTAGAGACAAAAAGTATGCGACATTTCTTGCGCAAAATGATTAAGTAATA 3978
 Db 440 IleProProArgArgGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 453
 QY 3979 AAATTTCATCAACAACTTAAT---TTAAAGAGCTTTCATCAATTCGACAGACAGAA 4035
 Db 454 LysIleMetThrAsnThrAsnGlnLysLysLysLysLysLysLysLysLysLysLysLys 473
 QY 4036 ACATCTCTCTCATGGTAT---TATATTAAGTAAGTGTGAAGAAATGATGATGAT 4092
 Db 474 ThrPheLeuLeuTrpGlnAsnGlnLysLysAspLysAsnGlnLysAlaGlnAspLeuAsp 493
 QY 4093 AAAGATTAAGAAAGCAAAATTCCTCCGATTTTGAATGATCCATGCTTACACATTT 4152
 Db 494 GluLysLysLysGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 513
 QY 4153 GGAGATTATAGAGATTTTATTGGAACAGATATA-----TCAAAA 4194
 Db 514 AlaAspTrpArgAspIleCysLeuGlyThrAspIleSerLysLysAspThrSerLys 533

QY 4195 GGTATGCTGAGGAGAACTAAACTTAAGACCAATAGATTCCTTTCAAAATGTGTAC 4254
 Db 534 GlyValGly-----LysValLysCysAsnIleAspAspValPhe----- 546
 QY 4255 CAAAATCTCTTAATGGA---AAAACAGCGCAAGAAATGGTGCAGACAAATGATGAG 4311
 Db 547 TyrLysIleSerAsnSerIleArgTrpArgLysSerTrpTrpGlnThrAsnIleProVal 566
 QY 4312 ATATGGAGACTTATGCTATGTGCATA-----GTAAAAAT 4347
 Db 567 IleTrpGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 586
 QY 4348 GGGGCAAAAAAGATGATTTTACCGAAACTACGTTACACACAGTCAAAATTT---AGT 4404
 Db 587 GluThrHisLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 606
 QY 4405 GACAAAACACCACTTTGGAGAAATTTGCCAAACGACCCAGTTTTCATGCTGATACC 4464
 Db 607 AspSerSerThrThrLeuSerLysPheSerGluArgProGlnPheLeuArgTrpLeuThr 626
 QY 4465 GAATGTCACGACATATTGCTATACAGACAAAAATTTGAAGAGTGTGCAGAAAA 4524
 Db 627 GluTrpGlnGlnAsnPheCysLysGlnGlnLysLysGluTrpLysLysLysLysLys 646
 QY 4525 TGTAACTCAATAGCAAAATGTAAGTGTATACAGATGTATAGAAA-----TGC 4575
 Db 647 CysLysAspCysAsp---ValAspGlyAspLysLysCysAsnGlyLysCysValAlaCys 665
 QY 4576 GAGGACTACGTTAA-----TATATGAAA 4599
 Db 666 LysAspGlnCysLysGlnIleThrHisSerTrpIleGlyIleTrpIleAspAsnTrpLys 685
 QY 4600 AAAAAAGAGATTCACAAAGATAA-----TATTAAGAGTGAACCGCAAA 4653
 Db 686 GlnLysGlyArgTrpGlnGlnLysLysLysLysLysLysLysLysLysLysLysLys 705
 QY 4654 AAAAGATTCGAT-----AACACACATTTGGTAAAGTGTACACATATCTGAAACG 4707
 Db 706 LysAsnSerAspAspAlaArgAspTrpLeuLysThrGlnLeuGlnAsnMetLysCysVal 725
 QY 4708 AATGCAACAGATTAATCTTAAC-----AGAAATTTCTCTGATGTT 4749
 Db 726 AsnGlyThrThrAspGlnAsnGlnCysGluTrpLysCysMetHisLysThrSerThrAsn 745
 QY 4750 GGTATTAAGCTGGAAGTGGCTGTGTGTCACAAAGAAATATCAATTTTACAAAAACAG 4809
 Db 746 SerAspMetProGluSer-----LeuAspGlnLys 755
 QY 4810 GCTTACTATGATCCGCAAAACATTTGGGTGCACAAAAATTTATGAAGATGACAGAAA 4869
 Db 756 ProGlnLysValLysAspLys---CysAsnCys-----ValProAsnGlnLysAsn 771
 QY 4870 TATACTAACATTTCCGCT----- 4887
 Db 772 AlaLeuSerValSerCysGlyPheProAspGlyGlnAlaPheGlyGlyValLeu 791
 QY 4888 AAAGATTAAGTGAAGAGATTAAGTGAAGGCAACAGCTGCTATTAACTGGCAAAAC 4947
 Db 792 GluGlyThrCysLysGlyLeu-----GlyGluProLys---LysLys 804
 QY 4948 AAAGCTCTAATTAATCAATTAATTTGAAGAAATGACTGAAG----- 4990
 Db 805 IleGluProGlnLysAspProThrAsnAspIle---LeuLysSerThrIleProValThr 824
 QY 4990 ----- 4990
 Db 824 rIleValIleuAlaLeuGlySerIleAlaPheLeuPheMetLysValIleTrpIleTrpVal 844
 QY 4990 ----- 4990
 Db 844 lTrpTrpIleTrpMetLeuCysValGlyAlaLeuAspThrTrpIleCysGlyCysIleCys 864

```
OY 4981 -ATGGCGTTTTCCTTGCCGCGAC-----TAGGTAAATTGTTTAAGCA 5034
      :::|||:: ||| |::::::::::
Db   864 silecysilhehellcycvalserValtyrvalcyssalftyrylvalyrAlpheleury 884
OY 5035 T-----TCGATGCCAATTATTCAGATCCACAGAAGTGAAGATAAAT 5076
      ::|||:: ||| |::::::::::
Db   884 tmetcysvalPherlyrlleyrpheelleryllertyrvalPhelleuleuylsmetlysly 904
OY 5077 GGGTTGCAGAAAAGATTGATGGAAGTGGCGGACAAGCATCAATTGGGTCAATAC 5136
      ::|||:: ||| |::::::::::
Db   904 S-----Metlysly 907
OY 5137 TACAAGAAAAAGAAAAAGAAAATATAAACGTGGATG 5179
      ||||| ||||| ::||| |||::|||::
Db   907 smetylslyslsmetylslyslsmetylslsalglyslslargile 921

RESULT 12
US-08-568-459A-4
Sequence 4, Application US/08568459A
Patent No. 5849306
```



```

Db      852  eserGlyVal-----AsnSerLysProLeuSerAspAspValArgProAs 867
Qy      6426  AAATTAAGAACTGAAATTAAGCGGTGCACATGATATAAAGCGCTTTTAAATGTC 6485
Db      867  pLySgLUleuGLuAspLnAsnSerAspGluSerGluGluThrValVal----- 884
Qy      6486  AGCAATAGAACTCAATTTTGTGGTTAAATATATATATGAAATTCGACAGAAAA 6545
Db      885  -----AsnHisIleSerLysSerProSerIleAsnAs 895
Qy      6546  TGAATTCGAAATGAAACAATTCACAGATGAAATTTAAAGAAATATATATATACATATG 6605
Db      895  nGlyAspAspSerGlySer-----GlySerAlaThrValSe 907
Qy      6606  TGAATTAAGATATGTTTGTGAACGTGATTTTCATATGATATAAATAATATATACGT 6665
Db      907  rGluSerSerSerSerAsnThrGlyLeuSerIleAspAspAspArgAsnGlyAspThr 927
Qy      6666  AACAAATAGTGTACACACATTCCTCAATGAAATTAATAGCAAAAAACAGATAAAAAA 6725
Db      927  eValArgThrGlnAspThrAlaAsnThrGluAspValIleArgGlyGluAsnAlaAsp 947
Qy      6726  AGATGAAGATTCGTAAATATTTTGGGAGAAAAATTAATTTATTTGGAGAGAAAT 6785
Db      947  sAspGluAspGluGlyGlyAlaAspGluGluArg----- 958
Qy      6786  GATATATGATTAATCATCATCTCACAGCAAGAAC-----GAAAAAGAAA 6833
Db      959  -----HisSerThrSerGluSerLeuSerProGluGluGly 972
Qy      6834  AATTAAGATTAATTAACAGTACATACATGACCAAACTGACGCTTCCCTGAGAGAT 6893
Db      972  tLeuThrAspAsnGluGlyGlyAsnSerLeuAsn-----HisGlu 986
Qy      6894  TGTAAAGAGCGCCCAATTTTGAATGGTTCACAGAAATGGCAGAAATTTGTGATAA 6953
Db      986  uValLysGluHisThrSerAsnSerAspAsnValGlnGlnSerGlyGlyLevalAsn 1006
Qy      6954  GAGGAAGCAACACTGTTTAA-----TTGAGCGCGG 6986
Db      1006  tAsnValGluGlyGluLeuLysAspThrLeuGluAsnProSerSerSerLeuAspGlu 1026
Qy      6987  CTGTAAAGAAATAGCTGATGATGATAT----- 7017
Db      1026  yLysAlaHisGluGluLeuSerGluProAsnLeuSerSerAspGlnAspMetSerAsn 1046
Qy      7018  -----GAGGTAAGACACAAAGATGTCAGAGCGCTGTATACATATCAAA 7064
Db      1046  rProGlyProLeuAspAsnThrSerGluGluThrGluArg-----IleSerAsnAs 1064
Qy      7065  TTTTATTAAGAACTGAAACTGAAATATGAAAGCAAAAGTTCACAAAGAT-- 7122
Db      1064  nGluTyLysValAsnGluArgGluAspGluArgThrLeuThrLysGluTyLysAsp 1084
Qy      7123  -----AAAGATGCCAAAGATTAAGATTAATCC 7151
Db      1084  eValLeuLysSerHisMetAsnArgLysSerAspArgGlyLeuTyLysArg----- 1102
Qy      7152  TTCTACTGAAGAGACATAGAGAAAGCAATGCTGCTCATGATATTTAAACATGAAT 7211
Db      1103  -----AsnSerAspLeuSerThrValAsnAspGluSerGluAspAlaGluAlaLys 1120
Qy      7212  AAAGAATTTATGCAAAATAGGATGTTCTTGTATGCAAAAACCTTTCACACATAC 7271
Db      1120  tLys-----GlyAsn-----AspThrSerGluMetSerHisAsnSerSerGlnHis 1136
Qy      7272  AAAACAACAACAACATCACAATCATCCATGCTAATGATATGCCAGATCGCTGAT-- 7329
Db      1136  eGluSerAspGlnGlnLys-----AsnAspMetLysThrValGlyLys 1151
Qy      7330  -----TATGTCCTGAAAGATTTAAACAGTGAAGTGTGCTGCTGCAACTTTCAAA-- 7377

```

```

Db      1151  uGlyThrThrHisValGlnAsnGluIleSerValProValThrGlyGluIleAspGlu 1171
Qy      7378  -----AAGGATCTATGATTCATACAAAAAATTAAGACCTAAATACC 7424
Db      1171  sLeuArgGluSerLysGluSerLysIleHis-----LysAlaGluGluGluArgLeu 1189
Qy      7425  TATGAATTCGTAGAGAAAGCAATATATTTATCTTAAAGACGAAATTAATATGGA 7484
Db      1189  rHisThrAspIleHisLys-----IleAsnProGluAspArgAsnSerAsnThr 1205
Qy      7485  TATTAACCTGAGCAAAAAATTTTATACCTATGAGTCTACAAAGCAAAAGCAATGAA 7544
Db      1205  rLeuHisLeuLysAsp-----IleArgAsnGluGluAsnGluArg 1218
Qy      7545  TAGTGTGACTAATTAATTAAT 7563
Db      1218  gHisLeuThrAsnGlnAsn 1224

RESULT 13
US-08-487-826B-4
; Sequence 4, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Slim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhuan
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 45
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbhe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487, 826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelien, Ned
; REGISTRATION NUMBER: 29, 655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-487-826B-4

Alignment Scores:
Pred. No.: 1,09e-38 Length: 1435
Score: 618.00 Matches: 339
Percent Similarity: 36.808 Conservative: 233
Best Local Similarity: 22.208 Mismatches: 538
Query Match: 3.18% Indels: 428

```


Db 566 glyleuValIglyIleAsnThrAsnSerIysThrValIHisAlaGlnLysLysAsn-- 585
 QY 5358 TACAGGAAATCCGGTAGTACTGCCGGAATAATTTCTGGAACGAAATAATACGAATGTCT 5417
 Db 586 -----AspLysleuPheArgAspIuTrpTrpLysValIleLysLysAspIa 601
 QY 5418 GTGGAACCATGATATCGGGGTACAAACGTGTGGGATGATGGAATATGGAATATG 5477
 Db 601 ITrpAsnValIleSerTrpValPheLys----- 610
 QY 5478 TGCAGAAAGTGAAGAATCTAAATAATGTGTTCTGTACTTCAGATGATGATTATCC 5537
 Db 611 -----AspLysThr 613
 QY 5538 TATGGGGAATAATCCGATGAGTACTGCTGATCAGTTTCTTCGATGGTTTCCGAAATG 5597
 Db 613 rValCysLysGlnAspAspIleGlnAsnIleProGlnPheArgTrpPheSerGluTr 633
 QY 5598 GGGTGAAGTTTTTTCGAACATTAAGAAAGAAATGGGAATTTGGTAGGGCGGTGA 5657
 Db 633 pClyAspAspTrpCysGlnAspLysThrLysMetIleGluTrpLysValGluCysLys 653
 QY 5658 TGATTTACTTGTGTGATTAATGAGATTAAGAAAGAAATGTACAGATCGCTGTACACA 5717
 Db 653 sGluLysProCysGlnAspAsp-----AsnCysLysSerLysCysAlaAsn 668
 QY 5718 ATATAAATAATTTATGCTAGTGAACACACAGTATGAAACAAATCAAAATAATG 5777
 Db 668 rTrpLysGluTrpLysSerLysLysGlnLysGlnLysGlnLysGlnLysGlnLys 688
 QY 5778 TGAG-----AATAAAGACAAATAATATCCGAGCATCTCTGTGGCAAAAGATCC 5825
 Db 688 nClyuTrpGlnLysGlnAsnAsnTrpLysMetTrpSerGluPheLysSerIleLysProG 708
 QY 5826 AGAGGACGCTCGCATATTTAGACAAACAAATTAATAAATAATTTGAAATAATAAGTG 5885
 Db 708 uValLysleuLysLysTrpSerGluLys-----CysSerAsnLeuAsn 723
 QY 5886 ACATTTGTAATTAAG-----TGT-- 5904
 Db 723 eGluAspGluPheLysGluGluLeuHisSerAspTrpLysAsnLysCysThrMetCysPr 743
 QY 5905 -----ATGAAGATGTGTCCACACACAGCATTAATCTGATTAATCAAAATATGCC 5960
 Db 743 oGluValLysAspValProIleSer--IleIleArgAsnAsnGluGlnTrpHisSerGln 762
 QY 5961 ATCATTAAGCATGAACAAAGAAAGTT-----GAAGAAAGTGAATTTGTC 6008
 Db 762 uAlaValProGluGlnAsnThrGluIleAlaHisArgThrGluTrpProSerIleSerG 782
 QY 6009 ACTGCCACGAGTCCACACACGTGTACGAAGGAACACCGTCACACGCGGTATCACTGAT 6068
 Db 782 uClyProLysGlnLysGlnGluGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 797
 QY 6069 ATCAAAAGCAGGAGTGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 6128
 Db 797 uSerLysIleSerValSer--ProGluAsnSerArgPro-- 809
 QY 6129 AAAAGTGAATAATCTAACACAGAAATGCGACACAAACACGACCCGACGACGACACA 6188
 Db 810 -----G 810
 QY 6189 ACAAAACAGAAACAGACATCAACAGCAACAACAACAGAAATGACGTGGC--ACAAT 6245
 Db 810 uThrAspAlaLysAspTrpSerAsnLeuLysLeuLysLysLysLysLysLysLysLys 830
 QY 6246 GGTAAAGCCATCTTTTTCGAATAAACAGATAGCAGGGGTGAATAAGAGGTGTGAATCC 6305
 Db 830 tProLysAlaValIleGlySerSerProAsn--AspAsnIleAsn-- 844
 QY 6306 AAAAAGCTATGACAAATCTTAATGGGGTGTATTTGAGTAAAGTCAAAAGAAATGA 6365
 Db 845 -----ValThrGluGlnGlnLysAspAsnIle 852

QY 6366 AATGGCATATGTATGCTCTAGAGAGAAAAAATTGATTAATTAATATCAATATTT 6425
 Db 852 eSerGlyVal-----AsnSerLysProLeuSerAspAspValArgProAs 867
 QY 6426 AATATATGAATGAAATTAACCGTCACATATATATAAGAGCGTTTATTAATATGTC 6485
 Db 867 pLysLysGlnLeuGlnAspGlnAsnSerAspGluSerGluGlnTrpValVal----- 884
 QY 6486 ACCAATAGAACTCAATTTTGTGTTAAATATATTAATTAATTAATTAATTAATTAAT 6545
 Db 885 -----AsnHisIleSerLysSerProSerIleAsnAs 895
 QY 6546 TGAATTCGAAATGGAACAAATTCAGATGAATTAAGAAATATGATTAATACATATG 6605
 Db 895 nClyAspAspSerGlySer-----GlySerAlaThrValSer 907
 QY 6606 TGATTTAAGATATGTTTTTGGACATATTTCTATATGATTAATAAATAATTAATAC 6665
 Db 907 rGluSerSerSerSerAsnThrGlyLeuSerIleAspAspAspArgAsnGlyAspThrPh 927
 QY 6666 AACAAATAGTGTAAACACCATCTCAATGAATAATATAAGAAACAGATAAAAA 6725
 Db 927 eValArgThrGlnAspTrpAlaAsnThrGlnAspValIleArgLysGlnAsnAlaAsp 947
 QY 6726 ACATGAAGATTAATCTTAATATTTTGGAGAAATAAATAATTTATTTGGAAAGAA 6785
 Db 947 sAspGlnAspGlnLysGlnLysAlaAspGlnLysGlnLysGlnLysGlnLysGlnLys 958
 QY 6786 GATATATGATTAATCAATATCAATCTCACAGACGAAAC-----GAAAAAGAAA 6833
 Db 959 -----HisSerThrSerGluSerLeuSerProGluGlnLys 972
 QY 6834 AATTAGATTAATTAACAGTACATGACATGACCAAACTGACGCTTCCCTGAAGAGTT 6893
 Db 972 tLeuThrAspAsnGlnGlnLysGlnLysSerLeuAsn-----HisGlu 986
 QY 6894 TGTAAAGGCCCCAATTTTTCAGATGTTTCACAGAAATGGCAAGAAATTTGTAAATA 6953
 Db 986 uValLysGlnLysIleThrSerAsnSerAspAsnValGlnGlnSerGlyLysIleValAsn 1006
 QY 6954 GAGGAAGAACAGTTGTTAAA-----TTGAGGCGGG 6986
 Db 1006 tAsnValGlnLysGlnLeuLysAspThrLeuGlnLysnProSerSerLeuAspGlu 1026
 QY 6987 CTGTAAAGCAATATGAGTGTAAATGCTAGTAAT----- 7017
 Db 1026 yLysAlaHisGlnGlnLysLeuSerGluProAsnLeuSerSerAspGlnAspMetSerAsn 1046
 QY 7018 -----GACGTAAGACACAGAAATGTGCAGAGCGGTGTGAACATATCAAAA 7064
 Db 1046 rProGluProLeuAspAsnThrSerGlnGlnLysThrGluArg-----IleSerAsnAs 1064
 QY 7065 TTTTATTAAGAGTGAAGAAACGAATATGAAACAAAGAAAGAAAGTTCACAAAGAT 7122
 Db 1064 nClyuTrpLysValAsnGlnArgGlnAspGluArgThrLeuThrLysGluTrpGluAsp 1084
 QY 7123 -----AAAGATGGCAAAAGTATAGGATATATCC 7151
 Db 1084 eValLeuLysSerHisMetAlaArgLysSerAspAspGluLysLeuTrpAspGlu----- 1102
 QY 7152 TTCTACTGAAGAGCATAGAGACAGCAACATGTGCTCATGAATATTAACATCAATAT 7211
 Db 1103 -----AsnSerAspLeuSerThrValAsnAspGlnSerGluAspAlaGluAlaLys 1120
 QY 7212 AAAAGATTAATATGGAATTAAGATTTGTTGTAACAAACCTTCTTCAACTATAC 7271
 Db 1120 tLys-----GlyAsn--AspThrSerGluMetSerHisAsnSerSerGlnHisIle 1136
 QY 7272 AAAAACAACACAAATCAATCAATCCATGCTTAATGATATGCAAGATCCCTGGAT 7329
 Db 1136 eGluSerAspGlnGlnLys-----AsnAspMetLysThrValGlyAspLe 1151

Query Match:	Best Local Similarity:	22.20%	Mismatches:	358
DB:	3.18%	4	Indels:	428
			Gaps:	72
US-10-087-013-1 (1-10628) x US-09-210-288-4 (1-1435)				
QY 3301	GAAGCTGTAAATGAATTAATGATTAATTAATGATTTATGGAAAGACAA-----TGC	3357		
DB 20	LysAlaArgAsnGluTyr-----AspIle-----LysGluAsnGluLysPheLeu	34		
QY 3352	AATTAATATCAGATTAATTAATCAAGAAATTAATCATGAACAGCAACAAATGCTGTAGTAAT	3411		
DB 35	AspAlaIleTyrLysGluLysPheAsnGluLeuAspLysLysTyrGlyLysAsnValGlnLys	54		
QY 3412	AGTGGTATGGAAGCTTCCAGTACTGCGCAAAATATCATATGACAGCAAGATTTTAATTT	3471		
DB 55	Thr-----AspLysLysIlePheThrPhe	62		
QY 3472	TTG---TCGAAATTAATTCACAAAAATGGTGGCAAAAGTAATTAATAAGT-----	3516		
DB 63	IleGluAsnLysLysLeuAspIleLeuAsnAsnSerLysPheAsnLysArgTyrLysSerTyr	82		
QY 3517	GGTACTAGTATGAAAGTGGCTGCTATGGTACTAATACACACGTATGAAAATGTTGGAGCA	3576		
DB 83	GlyThrProAspAsn-----IleAspLysAsnMetSerLeuIleAsn-----	96		
QY 3577	TATCTCCAGTAATACAGAAATTTTGATGATTCGTACGTACGACAAATAGCTTTGTGATGA	3636		
DB 97	--LysHisAsn--AsnGlu--GluMetPheAsnAsnAsnTyrGlnSerPheLeuSerThr	114		
QY 3637	AAAGATGATGGTATGAGATTAACGAAATAATGCTTTTAAGATTAACACACAGACCATGAT	3696		
DB 114	rSerSerIleIleLysGlnAsnLysTyrValProIle-----AsnAlaValArgVal	131		
QY 3697	GGTGGCTGTGGTGTGTAAGAAAGTGGATCGAACGACAGCAAGGTACAGATTAACGAAAAA	3756		
DB 131	IserArgIleLeuSerPheLeuAsp-----	139		
QY 3757	AAAGCGAAGAAAGATTAACGGAATGCTAAACAGTG-----AATGATTAATCTTAA	3806		
DB 140	-----SerArgIleAsnAsnGlyArgAsnThrSerSerAsnAsnGlnValLeuSer	156		
QY 3807	AGCA--AAAGATGGAAGAAAGAAACAGTAGAA--GATTGTCATCCAAAAAGAAATAGTA	3860		
DB 156	rAsnCySerGlnLysArgLysGlyMetLysTyrAspCysLysLysLysAsnAspArgSer	176		
QY 3861	TGGATATCCCGCATGTGGCAATGGGAAATATATAATTAGTGGAGACCTGTGTGTGAT	3920		
DB 176	rAsnTyr-----ValGlySerIle	181		
QY 3921	GGCCCGCTAGAGACAAAGATTAATGCGTACATTTCTGGCAAAATGATTAATGAATTAATAA	3980		
DB 181	eProAspArgIleGlnIleLeuLysIleValAsnLeuSer-----IleIleLys	197		
QY 3981	AATTAACATCAACAAGTATTAATAAAGAAAGCTTTTCATCAAAATTCGACAGCAACAAAT	4040		
DB 197	eThrTyrThrLysGluThrMetLysAspHisPheIleGluAlaSerLysLysGlu-----	215		
QY 4041	CTTCTCATGGTATTATTATAATAAGTAAGATGGTGAAGAAAGAACTGATGAATTAAGATT	4100		
DB 216	-----SerGlnLeuLeuLysLysAsnAspAsnLysTyrAsnSer-----	229		
QY 4101	AAAAAGAGCAAAATTCCTCCCGCATTTTGGAGATCCCATGTTCTACATATTTGGAGATTA	4160		
DB 230	-----LysPheCysAsnAspLeuLysAsnSerPheLeuAspLys	242		
QY 4161	TAGAGATTTTATTTTGGACAGATTAATATCAAAAGGTCATGGTGGAGGGAAGTAACATA	4220		
DB 242	rGlyHisLeuAlaMetLysAsnAspMetAspPheGly--GlyTyrSerThrLysAlaGln	261		
QY 4221	AGAGCAATATGATTTCTTTTTCAAAAAATGGTGACCAAAAATGCTCTAATGAGAAAACA--	4278		


```

Db 845 -----|||:::|||||
               -valThrgluIngluYaspasHII 852
OY 6366 AATGGCATATGTATGCTCCTAGAGAGAAAAAATTATGTAATAATATATACAAATATT 6425
      ::|||::|
Db 852 eSerGlyVal-----AsnSerLysProLeuSerAspValArgProAs 867
OY 6426 AATATGAACTGAAATAAAGCGTGACATGATATAAAGAGCGCTTATTAATATGTC 6485
      |||::|
Db 867 pLysLysgluLeuGluAspGlnAsnSerAspGluSerGluInlVal----- 884
OY 6486 AGCAATAGAACTCAATTTTGTGTTAAATATATATTAATTAATCTGCAGAGAAAA 6545
      |||::|
Db 885 -----AsnHisLysSerLysSerProSerLysAsnAs 895
OY 6546 TGAATTCGAAAAATGGACAAATCCAGATGATTTAAAGAAATATGATATATACATATG 6605
      ::|||::|
Db 895 nglYaspSerGlySer-----GlySerAlaThrValSe 907
OY 6606 TGATTTATTAAGATATGTTTTTGAACGATATTCTAATGATATAAAAAATTAATACGT 6665
      ::|||::|
OY 907 rglUserSerSerSerAsnThrGlyLeuSerLysAspAspArgAsnGlyAspThrPh 927
      ::|||::|
Db 927 eValArgThrGlnAspThrAlaAsnThrGluAspValLLeArgLysGluAsnAlaAspL 947
OY 6726 AGATGAAGATTAAGTAAATATTTTGGAGAAAAATTAATTTATTTGGAGAGAAAT 6785
      |||::|
Db 947 sAspGluAspGluLysGlyAlaAspGluGluArg----- 958
OY 6786 GATATATGATTAATCTATCATCTCACAGAGGAAAC-----GAAAAAGAAAA 6833
      |||::|
Db 959 -----HisSerThrSerGluSerProGluInlValSe 972
OY 6834 AATTAGAGATTAATACCACTGACATGACATGACCAAACTGACGCTTCCTCGACAGCT 6893
      |||::|
Db 972 tLeuThrGlnAsnGluGlyGlnAsnSerLeuAsn-----HisGluInl 986
OY 6894 TGTAAAAAGGCCCCAATTTTGTGATGGTTCACAGATGGCCAGAAATTTGTAAAT 6953
      |||::|
Db 986 vValLysGluHisThrSerAsnSerAspAsnValGlnGlnSerGlyGlyLeuAlaAsnSe 1006
OY 6954 GAGGAAGAACACTTGTAA-----TTGGAGAGCGGG 6986
      |||::|
Db 1006 tAsnValGluLysGluLeuLysAspThrLeuGlnAsnProSerSerSerLeuAspGlu 1026
OY 6987 CTGTAGGAATATGCTGTAATGCTAGTAAT----- 7017
      |||::|
Db 1026 yLysAlaHisGluGluLeuSerGluProAsnLeuSerSerAspGlnAspMetSerAsnTh 1046
OY 7018 -----GACGGTAAAGACAAAGATGTGCAGAGCGGTGTAAACATATCAAAA 7064
      |||::|
Db 1046 rProGlyProLeuAspAsnThrSerGluGluThrThrLysArg-----LLeSerAsnAs 1064
OY 7065 TTTTATTAAGATGGAATACTGATATGAAAGACAAAGAAAGTTCAAAAAGAT-- 7122
      |||::|
Db 1064 nclutYrLysValaAsnGluArgLysGluAspGluArgThrLeuThrLysGluYrGluAsp 1084
OY 7123 -----AAAGATGGCAAAAAGTAAAGTAAGATTTCC 7151
      |||::|
Db 1084 eValLeuLysSerHisMetAsnArgLysSerAspSerGlyLysLeuYrAspGlu----- 1102
OY 7152 TTTCTCTGAAAGACATAGAGAAAGCAACATGCTCATGATATATTTAAACATGAAT 7211
      |||::|
Db 1103 -----AsnSerAspLeuSerThrValaAsnAspGluSerGluAsnAspLysGluAlaLysSe 1120
OY 7212 AAAAGAAATTAATGATAGATGTTGCTGTATGCAAAAACCTTCTCACAACCTACC 7271
      |||::|
Db 1120 tLys-----GlyAsn--AspThrSerGluMetSerHisAsnSerSerGlnHisL 1136
OY 7272 AAAAACAACAACAACAATCATCATCGATGCTAATGATATGCCACAATCGCTGGAT-- 7329
      |||::|

```

```

Db 1136 eGluSerAspGluInlLys-----AsnAspMetLysThrValGlyAspLe 1151
OY 7330 -----TATGTCCTGGAAGATTTAAACAAGTGTGAGTCTCGAAGCTTCAAAA-- 7377
      ::|||::|
Db 1151 uGlyThrThrHisValGlnAsnGluLysSerValProValThrGlyGluLysAspGlu 1171
OY 7378 -----AAGGATCTATGATTCATCAAAAAAATTAAGTGAACCTAAATACC 7424
      |||::|
Db 1171 sLeuArgLysSerGlySerLysLysHis-----LysAlaGluGluGluArgLeuSe 1189
OY 7425 TATGATTTGTAGAGAAAGCAATATATTAATCTAATAAGAGCAAGAAATTAATGGA 7484
      |||::|
Db 1189 rHisThrAspLysHisLys-----LLeAsnProGluAspArgAsnSerAsnTh 1205
OY 7485 TATTACCTTGAAGAAAAATTTATACATGATGCTACAAAGAAAGAAAGAAATAAAA 7544
      ::|||::|
Db 1205 rLeuHisLeuLysAsp-----LLeArgAsnGluGluAsnGluArg 1218
OY 7545 TAGTGGACTATATTAAT 7563
Db 1218 gHisLeuThrAsnGlnAsn 1224

```

```

RESULT 15
US-08-568-459A-6
; Sequence 6, Application US/08568459A
; Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim I.
APPLICANT: Chiflis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-6

```

Alignment Scores: 1.39e-24 Length: 749
Pred. No.: 1

Score: 431.00 Matches: 210
Percent Similarity: 34.568 Conservative: 128
Best Local Similarity: 21.47% Mismatches: 262
Query Match: 2.22% Indels: 378
DB: 2 Gaps: 44

US-10-087-013-1 (1-10628) x US-08-568-459A-6 (1-749)

QY 7342 GAATTTACAACTGTGAGTGTCTGCAACTTTCAAAAAGGATCTATGATTCATACAAA 7401
|||||
DB 35 GluPhasInlySile-----PheProGluArgInValGlnIleHisIleSer 50
QY 7402 AAAATACGACCTAAATACCTATGATGTGTAGAGAAACAGCATTTATTATCT 7461
|||||
DB 51 AsnIle-----PheLys 54
QY 7462 AAAGAAGCAAAATATATATGATATT-----ACCTTGAGGAAAATTTATACCT 7512
|||
DB 55 GluTyrLysGlnuSnsValAspIleIlePheGlyThrIeuSntYrGluTyr----- 72
7513 ATTGAGCTTACAAAGAAAAGAAAGTAAATAGTTGACATAATAATATCTTGGCAT 7572
|||||
DB 73 AsnAspPheCysLysGlnuSnsValSerProGluLeuValSerAlaIleLysTyrAsn----- 89
QY 7573 CCTAAGAAACCTTATGCACCTGATTAATATATGAGAAAGAAACCTTGTGAATATGA 7632
|||||
DB 90 LeuLysAlaProAsnAlaLysSerProArgIleTyrLysSerLys-----GluHisGlu 107
QY 7633 GAAGAATAATCGTTT-----AGGTAGATTTATGAATGAAATGT 7671
|||
DB 108 GluSerSerValPheGlyCysLysThrLysIleSerLysValLysLysLysStrpAsnGly 127
QY 7672 TACAAAATTCCAAGATTCATACAGAGAAAAGATATGTGTACCTCCAGAGAGAA 7731
|||||
DB 128 TyrSerAsnAsnLys---ValThrLysProGluGlyValCysGlyProProArgArgGln 146
QY 7732 CATATGCTGTAGATTTAGATGAATTTAA-----ATTGAAGACTTAAGAT 7782
|||||
DB 147 GlnLeuCysLeuGlyTyrIlePheLeuIleLeuArgAspGlyAsnGluGluLysLeuLysAsp 166
QY 7783 AGTAATATCTCTAAAAATGTTCTGTCGACAGCAAGTAATGAGAAATGACATATA 7842
|||||
DB 167 His-----IleAsnLysAlaIleAsnTyrGluAlaMetHisLeuLys 180
QY 7843 AAAAATCTCAATCAGAGAAAGCGGTGCGCAATATCAATATGTATGATTAATATAT 7902
|||||
DB 181 GluLysTyrGlnuSnsValGlyGly-----AspLysIleCysAsnAlaIleLeuGly 197
7903 AGTTGCGTGTGCTGTCGTCATGATAGTAGAGCAAGCATATGTACCAATTTGGTGTAC 7962
|||||
DB 198 SerTyrAlaAspIleGlyAspIleValArgGlyLeuAspValTrrpArg----- 213
QY 7963 TTACTCCCGCTAGAAATA-----AAATTTATTAAGTTTGTGATCATATATGA 8013
|||||
DB 214 -----AspIleAsnThrAsnLysLeuSerGlnuLysPheGlnLysIlePheMet 229
QY 8014 AAATGAGAAATATAAATAGGTAGAAATATAACAGCATGTACAAAGCTTTCCTCT 8073
|||||
DB 230 GlyGlyGlyAsnSerArgLysGlnuSnsValAspAsnAsnGlnu-----ArgAsn 245
QY 8074 GCTTGTGGATCTGATAGAAAGATATTTGCAAGCATG----- 8115
|||||
DB 246 LysTrrpIlePheLysGlnuArgAsnLeuIleTrrpSerSerMetValLysHisIleProLys 265
QY 8116 -----ACGTGCAAGCACACAGAGATGCAAAACTTTTGAAGAGAGAAATGATGGA 8169
|||||
DB 266 GlyLysThrCysLys-----ArgHisAsnAsn 274
QY 8170 TTTGAACGCAATTAATATACAAAGTAAGTGTGACATAGAGACGATCCACCTGTTGAT 8229
|||||
DB 275 PheGluLys----- 277
QY 8230 GATTTATATACCTCAACGCTTTCATGATGATGATGATGCTGCTGAATTTATTTGTAAGCA 8289

DB 278 -----IleProGlnPheLeuArgTrrpLeuLysGlnuTrrpLysGlnuPheCysGlnu 295
QY 8290 CTGATGCAAAATTTGCAAAATTTAAATCATGTATCATCTGTAATCAATCATGACAGCA 8349
|||||
DB 296 MetCylThrGlnuValLysGlnuLeuGlnuLysIleCysGlnuSnsValAsnCysSerGlnuLys 315
QY 8350 TGCAGAGATGATTTATGATGAAATTAAGTGTGACAGCTGTAAACGAGATGTCAAGATAT 8409
|||||
DB 316 -----LysCysLysAsnAlaCysSerSerTyr 324
QY 8410 AAAAATTTGCTCTTAATGGAATCTCTATGTGATATACAAATCAATTAATACAAAGAA 8469
|||||
DB 325 GluLysTrrpIleLysGlnuArgLysAsnGlnuTrrpAsnLeuIleSerLysLysPheAspSer 344
QY 8470 TTGTATGACAAACCAATATATATACAAATCTCTACTATGATCATGTTCAAAATTTGTGA 8529
|||||
DB 345 -----AspLysLysLeuAsnLysLysAsnAsnLeuTyrAsn----- 356
QY 8530 CAAAAGTTGAAACCTTTAAAGTAAATGTTCTGTGAGAGCTTTTCTGATATCTTCAT 8589
|||||
DB 357 -----LysPheGlnuSpsSerLysAlaTyrLeuArg 366
QY 8590 GAAACAGTAG---TGTGATATTAATTTATGATGAAATGATGCTTCTTCCAAATATA 8646
|||||
DB 367 SerGluSerLysGlnuCysSerAsnIleGluPheAsnAsp----- 379
QY 8647 CGAATATGCTTTGCAAGAAACACCAAAAGTATTAAGAGCTTGCAGT---TGTACA 8703
|||||
DB 380 GluThrPheThrPhe-----ProAsnLysTyrLysGlnuAlaCysMetValCysGlnu 396
QY 8704 CTACCTCTTCAAGAT-----CCATGTGATTAAT-----TGTCTTACGATCAAA 8745
|||||
DB 397 AsnProSerSerSerLysAlaLeuLysProIleLysThrAsnValPheProIleGlnu 416
QY 8746 AACCAATATGATGATGATGATTAACAACTTTTACCTTCTGCTGGAAGATGATTAATGAT 8805
|||||
DB 417 SerLysLys-----SerGluLeuSerSerLeuThrAspLysSerLysAsn----- 431
QY 8806 AATTAATCTGATTAATGGAAGCATACCTGTTCTTATGTTGATGATGATTAACAAAGT 8865
|||||
DB 432 -----ThrProAsnSerSerGlyGly----- 439
QY 8866 GTATGATTCCTCCAGAGAGACATTTATGTAACAAGCTATCACTGATATATATAT 8925
|||||
DB 440 -----AsnTyr 441
QY 8926 AGAAAGCTGATAAGAAATTTTAAAAA-----AAACTT 8961
|||||
DB 442 -----GlyAspArgGlnIleSerLysArgAspValHisHisAspGlyProLysGlnu 459
QY 8962 CTACTCTGCTTCTCACTCAAGGACATTTGTAGTCAAAAATTAATATGCGAAGAGAG 9021
|||||
DB 460 ValLysSerGlnuLysGlnuValProLysIleAspAlaValLysThrGlnuAsnGlnu 479
QY 9022 TTGTGC-----TTGAGGCAATGAATAATAGTATGACATTAATTCGAT 9066
|||||
DB 480 PheThrSerAsnArgAsnAspIleGlnuLysGlnuLysSerLysGlnuAspHisSerSer 499
QY 9067 ATATTTAAAGCACTGATATGATGAGACATTCATATCTGAAAAATTAATAAATTTT 9126
|||||
DB 500 ProValHisSerLysAspIle-----LysAsnGlnu 509
QY 9127 GAAACATCAATATGAGCAACGCAAAATCGTAACATGATGCGGAAATAATATAGACGTGAC 9186
|||||
DB 510 GluProGlnuArgValLysSerGlnuAsn----- 518
QY 9187 ATATGCAAGCTATGTTATGTGATATATAAATGCTACTTCAAAAGTAACATAGATGAA 9246
|||||
DB 518 ----- 518
QY 9247 GATGTGTCAATTTACCAAGATGAGAAACATATCATGTTTCTTGTGTTAATTGAA 9306
|||||

```
Db 519 -----LeuProLysIleGlu----- 523
OY 9307 TGGGCAAGCAAGCATGTAAGAAACATGTAAGTTCATTAACAAATGT 9366
Db 524 -----GluLysMetGluSerSerIleProIle----- 534
OY 9367 CCTCGTCAAGCAAGATTAATTTTGAAGCGTCAGAAATTAAGCAACCTGGATGTCAG 9426
Db 535 -----ThriHisIleGluIleGlu----- 540
OY 9427 AATGATATTAGAAAATATATTAGCTGTAATATATGATATAAAATACATGCAAAATCTA 9486
Db 540 ----- 540
OY 9487 AATATTAATATTAAGCAATTAAGATCAATCTTCAGTAATATAGACAAATACCATCT 9546
Db 541 -----LysGluGlnSerSerAsnSerSerAspAsnProIle 553
OY 9547 -----GAGAA 9552
OY 554 ValValSerGlyArgGluSerLysAspValAsnLeuHisThrSerGluArgIleLysGlu 573
OY 9553 AATGTCAGTCAATATTAATCAAAAGATCTCAATGCCCTTGGAGTAA----- 9603
Db 574 AsnGluGluGlyValIleLysThrAspSerSerLysSerIleGluIleSerLysIle 593
OY 9604 -----AATGATATTAATGAATAGTACAGAAACAAAAATTAAT 9642
Db 594 ProSerAspGlnAsnAsnHisSerAspLeuSerGlnAsnAlaAsnGluAspSerAsnGln 613
OY 9643 GAAATTAATGAA-----TTCAAAAGAGTACTAAAAATTAATATCT 9684
Db 614 GlyAsnLysGluThrIleAsnProSerThrGluLysAsnLeuLysGluIleHis--- 632
OY 9685 GATTATATTTTGTGAGATGCAACACAAAAATCATGTACTAGATGAATATAAA 9744
Db 633 -----TyrLysThrSerAspSerAspHisGlySer 643
OY 9745 GAAGAAGCAAAACAGTTCGTCCTAAGCAGCTATTTCTTTACACCCCATGTAGATTCT 9804
Db 644 LysIleLysSerGluIleGluProLysGluLeuThrGluLysProLeuThrAspLys 663
OY 9805 TTCTATCAAGCAGCTTTATTTCTCAACACATGAGTACACAAATATGATCTAAAAATGAT 9864
Db 664 -----LysThrGluSerAlaAlaIleGlyAspLysAsnHisGlu 676
OY 9865 ATATTGAAAGTAGT----- 9879
Db 677 SerValLysSerAlaAspIlePheGlnSerGluIleHisAsnSerAspAsnArgAspArg 696
OY 9880 ---ATCTCTGTGTATTGTA-----TCGCGCTTAGCTTTGATACCGCTT 9921
Db 697 IleValSerGluSerValValGlnAspSerSerGlySerSerMetSerThrGluSerIle 716
OY 9922 CATTTCATGAAGAAAAATTCAAATCGTCTGCGACTGTGCGTATCTGAAT 9975
Db 717 ArgThrAspAsnLysAspPheLysThrSerGluAspIleAlaProSerIleAsn 734
```

Search completed: April 28, 2003, 12:04:25
Job time : 577.5 secs

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: April 28, 2003, 11:01:12 ; Search time 653.5 Seconds
(without alignments)
6701.970 Million cell updates/sec

Title: US-10-087-013-1
Perfect score: 1945
Sequence: 1 atggggcttcctgcgaata.....atatatgatgaagtatg 10628

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Archived: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV-x1p
-Q/cgn2.1/USPTO.spool/US10087013/runat.28042003.102845.20267/3pp.query.fasta_1.10823
-DB-spprembl.21 -OEMT-fastan -SUFFIX-n2p.rspt -MIMARCH-0.1 -LOOPL-0
-LOOEXT-0 -UNITS-bits -START-1 -END-1 -MATRIX-biosum62 -TRANS-human40.cdi
-LIST-45 -DOCALLIGN-200 -THR_SCORE-pct -THR_MAX-100 -THR_MIN-0 -ALIGN-15
-MODE-LOCAL -OUTFN-plt -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-2000000000
-USER-US10087013/cgn.1.1115/runat.28042003.102845.20267 -NCPU-6 -ICPU-3
-NO_XLPHY -NO_MAP -LARGEQUERY -NEG_SCORES-0 -WAIT -LONGLOG -DEV TIMEOUT-120
-WARN TIMEOUT-30 -THREDS-1 -XGAPOP-10 -XGAPEXT-0.5 -FGAPOP-6 -FGAPEXT-7
-YAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database: SPTREMBL.21:*

- 1: sp-archaea:*
- 2: sp-bacteria:*
- 3: sp-fungi:*
- 4: sp-human:*
- 5: sp-invertebrate:*
- 6: sp-mammal:*
- 7: sp-mhc:*
- 8: sp-organelle:*
- 9: sp-phage:*
- 10: sp-plant:*
- 11: sp-rodent:*
- 12: sp-virus:*
- 13: sp-vertebrate:*
- 14: sp-unclassified:*
- 15: sp-ribose:*
- 16: sp-bacteriophage:*
- 17: sp-archae:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	19407	99.8	3542	5	Q905M2	Q905M2 plasmodium

2	12954.5	66.6	3287	5	Q8T326	Q8T326 plasmodium
3	7367	37.9	1615	5	Q8T325	Q8T325 plasmodium
4	4515.5	23.2	2706	5	Q15870	Q15870 plasmodium
5	4260	21.9	3006	5	Q26032	Q26032 plasmodium
6	4010	20.6	3026	5	Q26030	Q26030 plasmodium
7	3501	18.0	3078	5	Q26031	Q26031 plasmodium
8	3328.5	17.1	2527	5	Q95M83	Q95M83 plasmodium
9	3154	16.2	2924	5	Q25733	Q25733 plasmodium
10	2932	15.1	1685	5	Q9U4A2	Q9U4A2 plasmodium
11	2859	14.7	2658	5	Q8T5G0	Q8T5G0 plasmodium
12	2817	14.5	510	5	Q8T6L0	Q8T6L0 plasmodium
13	2727	14.0	2647	5	P90580	P90580 plasmodium
14	2726.5	14.0	2664	5	Q26033	Q26033 plasmodium
15	2678.5	13.8	2212	5	Q94657	Q94657 plasmodium
16	2556.5	13.1	494	5	Q8T6K7	Q8T6K7 plasmodium
17	2480.5	12.7	492	5	Q8T6K9	Q8T6K9 plasmodium
18	2457	12.6	2710	5	Q9XZB8	Q9XZB8 plasmodium
19	2449.5	12.6	494	5	Q8T6K0	Q8T6K0 plasmodium
20	2441.5	12.5	465	5	Q8T6K4	Q8T6K4 plasmodium
21	2413.5	12.4	494	5	Q8T6K5	Q8T6K5 plasmodium
22	2412.5	12.4	492	5	Q8T6K6	Q8T6K6 plasmodium
23	2403.5	12.4	496	5	Q8T6K8	Q8T6K8 plasmodium
24	2335.5	12.0	461	5	Q8T6K3	Q8T6K3 plasmodium
25	2303	11.8	2135	5	Q61077	Q61077 plasmodium
26	2258.5	11.6	460	5	Q8T6K1	Q8T6K1 plasmodium
27	2076.5	10.7	1327	5	Q9NFB4	Q9NFB4 plasmodium
28	2055.5	10.6	427	5	Q8T6K2	Q8T6K2 plasmodium
29	1890.5	9.7	2169	5	Q97312	Q97312 plasmodium
30	1773	9.1	2197	5	Q96296	Q96296 plasmodium
31	1755	9.0	2163	5	Q9NFB6	Q9NFB6 plasmodium
32	1755	9.0	2182	5	Q26034	Q26034 plasmodium
33	1744	9.0	2209	5	Q97324	Q97324 plasmodium
34	1731.5	8.9	2228	5	Q60991	Q60991 plasmodium
35	1709.5	8.8	2277	5	Q9U0G5	Q9U0G5 plasmodium
36	1614	8.3	2209	5	Q9U0G6	Q9U0G6 plasmodium
37	1586	8.2	1711	5	Q96108	Q96108 plasmodium
38	1542.5	7.9	2042	5	Q25766	Q25766 plasmodium
39	1249	6.4	1729	5	Q25734	Q25734 plasmodium
40	924	4.7	173	5	Q96735	Q96735 plasmodium
41	835.5	4.3	431	5	Q96294	Q96294 plasmodium
42	750.5	3.9	2647	5	Q9U4X0	Q9U4X0 plasmodium
43	665.5	3.4	438	5	Q9Y1N7	Q9Y1N7 plasmodium
44	658.5	3.4	455	5	Q9Y1N6	Q9Y1N6 plasmodium
45	651	3.3	118	5	Q964S0	Q964S0 plasmodium

ALIGNMENTS

RESULT	ID	Q905M2	PRELIMINARY	PRT	3542 AA.
AC	Q905M2				
DT	01-MAY-2000	(TREMREL. 13, last sequence update)			
DT	01-MAY-2000	(TREMREL. 13, last sequence update)			
DT	01-MAY-2002	(TREMREL. 20, last annotation update)			
DE	FCR3	CSA ligand (Fragment).			
GN	VAR.				
OS	Plasmodium falciparum.				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxID=5833;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=FCR3;				
RX	MEDLINE=20006305; PubMed=10535993;				
RA	Buffet P., Gamain B., Scheidig C., Baruch B., Oishi S., Fujii N.,				
RA	Fusai T., Parzy D., Miller L.H., Gysin J., Scherf A.,				
RT	"Plasmodium falciparum domain mediating adhesion to chondroitin				
RL	sulfate A: A receptor for human placental infection."				
EMBL	AF133811; CAB59840.1; -				
DR	InterPro: IPR001219; Neurotoxin.				
DR	InterPro: IPR004258; PFEKP.				
DR	Pfam: PF03011; PFEKP; 1.				

DR PRINTS: PR00284; TOXIN.
FT NON TER 3542 3542
SQ SEQUENCE 3542 AA; 413089 MM; 97085EE88DA2EC2 CRC64;
Alignment Scores:
Score: 0 Length: 3542
Percent Similarity: 19407.00 Matches: 3542
Best Local Similarity: 100.00% Conservative: 0
Query Match: 99.75% Mismatches: 0
DB: 5 Indels: 0 Gaps: 0
US-10-087-013-1 (1-10628) x Q9U5M2 (1-3542)
QY 1 ATGGGGTCTCTCGCAATATTTATTTAAATGGGGAATGCACATCATCTATAG 60
DB 1 MetGlyPheSerCysLysTyrPheIleIleLysMetGlyAsnAlaIleSerLeuGlu 20
QY 61 CGAGATGCTAAAGCCCTATTATTAAGAAAGTCACAAAGTCGACAGAAATGTTGGAA 120
DB 21 GlyAspAlaLysSerProIleIleLysGlySerIleLysSerAlaIleGlnValLeuGlu 40
QY 121 CGTATGCGCAAAATATTAAGCATCCATCAAAATATGCAAAAGACATGTGATGCTTG 180
DB 41 ArgTyrAlaLysAsnIleArgHisProSerLysTyrAlaLysGlyHisValAspSerLeu 60
QY 181 AAAGGGATTTGAGCAAGACGAATTCGTGGTGCCTTCTAGCCAGCAATTAAGCAT 240
DB 61 LysGlyLysPheLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 80
QY 241 AATTATATATTCATTCATTCATGTAATTAAGATCAATAGACATCTAATTAAGCAT 300
DB 81 AsnTyrTyrTyrProTyrProCysAsnLeuAspHisLysGlyHisIleHisLeuArgTyr 100
QY 301 GATGATGTGAATTTGAGACATCTTGCATGTAAGCAACCAACCGATTTGATGAAGAT 360
DB 101 AspAspAlaLysLeuArgHisProCysHisLysGlyArgGlyGlnAsnArgPheAspGluAsp 120
QY 361 GAAGAATTCGAATGTGAAATTAATACGTAAATTAAGAAAGAAATGATGCTATAGCC 420
DB 121 GlnGlnSerGlnCysGlnLysAsnLysLysLysLysLysLysLysLysLysLysLys 140
QY 421 TGTGCCACCTGAGAGACGACATGATGTAATTAAGAAAGCTTGAAGCTTAAATGATA 480
DB 141 CysAlaProProArgArgArgHisPheCysAspLysAsnLeuGlnAlaLeuAsnAspIle 160
QY 481 AATACCAAAATATTCATGATTAATGGAATGTACTAGTACGCAAAATATACGAAGT 540
DB 161 AsnThrGlnAsnIleHisAspLeuLeuGlnValLeuValThrAlaLysTyrGlnGly 180
QY 541 GAATCAATGTAAATATCATCCACATTAAGAACTTCAAGCGCTGTGCTCTTGA 600
DB 181 GluSerIleValAsnAsnHisProHisLysGlyThrSerAspAlaCysThrAlaLeuAla 200
QY 601 CGAAGTTTGCAGATATAGTGATATTTGAAGAGAAATAGATGTTAAACCAAGTGC 660
DB 201 ArgSerPheAlaAspIleGlyAspIleValArgGlyIleAspPheLysProAsnVal 220
QY 661 CATGCAAGTAAAGAAAGCGGTCTCCGACAGGTTTCAAGAAATATACATGATGAAGAA 720
DB 221 HisAspLysValGlnThrGlyLeuArgGlnValPheLysLysIleHisAspGlyMetGln 240
QY 721 GATCAAGTAAAGAAATGATTAATCTCGATGATGATCTGAAATTTATTAATTAAGAA 780
DB 241 AspGlnValLysAsnAspTyrAsnProAspLysSerGlyAsnTyrTyrLysLeuArgGln 260
QY 781 GCATGTGGAATGTGAATAGAAATAGATGGAACCTTAACATGATGATGATGAT 840
DB 261 AlaTyrPheAsnValAsnArgAsnLysValThrPheLysIleHisPheCysAspAlaSerTyr 280
QY 841 AAATCTGGAATTTATTAAGCAATGCAAGAAATATACCATTAATTTCTAATGCG 900
DB 281 LysSerIleTyrPheMetGlnSerGlnSerAsnThrProLeuPheSerAsnProLysCys 300

QY 901 GCCCATTAACAGAAAGTTCTTACCATTATGATTAATGCTCCCTCAATATTTAGCTGG 960
DB 301 GlyHisLysGlnLysValProThrAsnLeuAspTyrValProGlnTyrLeuArgTyr 320
QY 961 TTCAGCAATGGGAGAGAGTTTGGCCAAAAGAAATATTAATTAAGTAAAGTCAAG 1020
DB 321 PheAspGlnTyrPheLysGlnLysLysLysLysLysLysLysLysLysLysLysLys 340
QY 1021 GACTCCTGTGTAATGACAAAGACCTTATATGATGATGATGATGATGATGATGATG 1080
DB 341 AspSerCysArgAsnAspLysGlnArgLeuTyrCysSerHisAsnGlyHisAspCysThr 360
QY 1081 ACACTATTGGAAAAAGGATTTTGGCATTTGCAATATAGTCACTGATGCACT 1140
DB 361 ThrThrIleTyrPheLysGlyLysLysLysLysLysLysLysLysLysLysLysLys 380
QY 1141 AAATGCAAGTTTGAAGTTGTTAGGAATCAACAAGACATTAAGAAAGCAATTAAG 1200
DB 381 LysCysLysValPheGlnValThrPheLysGlnGlnGlnGlnGlnGlnGlnGlnGln 400
QY 1201 GAAAAATATGAAAGAAATATACATCATATTTATGCAAGATTAACAAATTTGCAAT 1260
DB 401 GlnLysTyrGlnLysGlnLysLysLysLysLysLysLysLysLysLysLysLys 420
QY 1261 AATTATAGTGAATATTAATTAACAAATTTATGCAAGATTAAGAAAGCAATATGCA 1320
DB 421 IleAsnSerGlnTyrTyrLysGlnPheTyrGlnLysLysLysLysLysLysLysLys 440
QY 1321 AATGCACTTTTAAATTAATTAATGAAGAAAGATTTGAAGAGATTAATGCAAG 1380
DB 441 AsnAspThrPheLeuAsnLeuLeuAsnGlnLysLysTyrCysLysGlyLeuProGly 460
QY 1381 GAAAAATATTAATTAATTAATTAATGAAGATTAATGAAGATTAATTAATGCTTGA 1440
DB 461 GlnLysAspIleThrPheThrAsnSerAlaAspAspLysGlyIlePheTyrArgSerGln 480
QY 1441 TATTGCCAAGTGTCTCCGACGTCGGGGTCAAAATGATGATGATTAATTAATGCA 1500
DB 481 TyrCysGlnValCysProAspCysGlyValLysCysAspLysLysLysLysLysLys 500
QY 1501 TCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
DB 501 SerAspAsnAspArgGlnArgValAsnAsnGlnLysPyrLysProProTyrGlyValLys 520
QY 1561 CTTACTAATATCACTGCTTATATGATGATGATGATGATGATGATGATGATGATGAT 1620
DB 521 ProThrAsnIleThrValLeuTyrSerGlyAsnGlnGlnGlyAspIleThrGlnLysLeu 540
QY 1621 GAAAAATTTTGAACAGCTCACTAATTAACAAAGATTAATTAATTAATGCAAGTGC 1680
DB 541 GlnAsnPheCysAsnSerSerThrAsnTyrLysAspLysAsnAsnGlnLysTyrPheLys 560
QY 1681 TATTATAGGATGAATTAATTAATTAATTAATGAAGTGAAGCAAAATTAATTAAT 1740
DB 561 TyrTyrLysAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 580
QY 1741 GATTAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1800
DB 581 AspAspProLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 600
QY 1801 AGGATACATTAATGAAGGAAATGACAAACTTAATCTGATTAATTAATTAACAC 1860
DB 601 ArgAspThrIleLysTyrAsnAspLysLysLysLysLysLysLysLysLysLysLys 620
QY 1861 TGTATGATGAATGTAAAGAAATGCTATGTTTGAACATGCTTGAACAAAGAA 1920
DB 621 CysIleAspGlnCysAsnArgAsnGlnCysLysLysLysLysLysLysLysLysLys 640
QY 1921 GAAGAATGGAATAGTAAAGAACTGTCACAAAAAGAAATATTAACAGCAATGCTAT 1980
DB 641 GlnGlnTyrPheSerIleLysLysLeuPheThrLysLysLysLysLysLysLysLysLys 660

Db 1381 PheTyrThrPheGlyAspTyrAlaAspPheLeuPheGlyThrAspIleSerLysGlyHis 1400
QY 4201 GGTGAGGAGATGTAACCTAAAGACAAATAGATCTCTTTCAAAATAGTGCACCAAAA 4260
Db 1401 GlyGlnGlySerLysLeuLysGlnIleAspSerLeuPheLysAsnGlyAspGlnLys 1420
QY 4261 TCTCCTAATGGAAGAAACGCCAAGAAAGGTGACAGAACATAGTCAATATGGA 4320
Db 1421 SerProAsnGlyLysThrArgGlnIleTyrPheGlnHisSerHisGlnIleTyrGln 1440
QY 4321 GCTATGCTATGTCAGTATGTAATAATGGGCAAAAAGATGATTTACCGCAAACTAC 4380
Db 1441 AlamelLeuCysAlaLeuValLysIleGlyAlaLysLysAspSerPheThrGlnAsnTyr 1460
QY 4381 GGTTCACAACAGCTCAAAATTTAGTACAAAAGCACCACTTTGAGGAATTTGCCAAA 4440
Db 1461 GlyTyrAsnAsnValLysPheSerAspLysSerThrThrLeuGlnGlnPheAlaLysArg 1480
QY 4441 CCCCAGTTTTCAGTATGCTAAGCAATGTAGCGACGCTATGCTATACGCAAAA 4500
Db 1481 ProGlnPheLeuArgTyrPheThrGlnIleTyrPheAspPheTyrCysTyrThrArgGlnLys 1500
QY 4501 TATTTGAAGCATGTGCAGGAAATGTAACTCAATGACCAATGAGTGTGATACAGAA 4560
Db 1501 TyrLeuLysAspValGlnGlnLysCysLysSerAsnAspGlnLeuLysCysAspThrGln 1520
QY 4561 TGTAAATGAAGATGCGAGGACTACGTTAAATATATGAAAAAAGAGTGCATGCCA 4620
Db 1521 CysAsnLysLysCysGlnAspTyrValLysTyrMetLysLysLysGlnTyrIlePro 1540
QY 4621 CAAGATTAATTTACAAAGATGAACGCGACAAAAAAGATTCGATAGCAACATTTGGT 4680
Db 1541 GlnAspLysTyrTyrLysAspGlnArgAspLysLysArgPheAsnArgGlnHisIleGly 1560
QY 4681 GTATGGTTCAGACTTACTGAGACGATGCAACAGATTTACTTGAACGAAATTTACT 4740
Db 1561 ValMetValThrAspTyrThrGlnTyrThrAsnAlaThrAspTyrLeuAsnArgLysPheThr 1580
QY 4741 GCTAGTTGGTATGAAGCCTGGAAGTGCCTGCTGCTCAAAAGAAATATACAAATTTGTA 4800
Db 1581 AlaSerCysGlnLysAspLysProGlnLysSerAlaSerValGlnIleArgAsnIleGlnLeu 1600
QY 4801 GAAAAACAGGCTTACTATGATGCGCAACAAACATTTGGTGCACAAATTTATTTGAAAT 4860
Db 1601 GlnLysGlnAlaLysTyrTyrAspAlaAspLysHisCysGlnCysThrLysPheIleGlnAsn 1620
QY 4861 GACGACAAATATATCTAATCACTTTCGATTAAGATAGTGCAGAGATTTAGTAAAGAGCA 4920
Db 1621 AspAspLysTyrThrAsnIleSerSerLysAspLysCysLysGlyLeuValLysGlnAla 1640
QY 4921 AACACAGGTGCTATTAAGTGGCAAAACAAAGTCTTAATATACATTAATTTGAAAGA 4980
Db 1641 AsnThrGlnAlaIleLysTyrPheLysAsnLysGlnProAsnAsnTyrAsnAsnLeuLysGln 1660
QY 4981 TTGACGAGAGATGTGCTTTCTCTCTGCTGCTACGATATATGTTTCATGATGATGAT 5040
Db 1661 LeuThrGlnLysValLeuPheProSerArgArgLeuArgLysPheHisAlaLeuAsp 1680
QY 5041 GCGCAATATACGATCCAGAACTTAAGATGAAATGGTTCGCAAAAAGATTTGATGGA 5100
Db 1681 GlnAsnTyrThrAspProGlnValLysAspGlnAsnGlyLeuAlaGlyLysArgLeuMetGln 1700
QY 5101 GTGGCGGAGAGGATGATCAATTTGGGCTCAATATCTAATGAGGCTCCCGCTTGATGCTATG 5160
Db 1701 ValAlaAlaThrGlnGlnTyrAsnLeuGlnIleTyrTyrLysGlnLysGlnLysGln 1720
QY 5161 AAAATTAAGAGCTGCATGACCAAAATATCTTATGAGGCTCCCGCTTGATGCTATG 5220
Db 1721 LysIleLysThrSerAspAlaHisLysTyrSerTyrGlnValProProCysSerAlaMet 1740
QY 5221 AAATATAGTTTATGATTAAGAGATATAATCTAGGTAATTTGATTTGGAAGATGAA 5280
Db 1741 LysTyrSerPheTyrAspLeuArgAspIleIleLeuGlnIleAspAsnLeuGlnAspGln 1760

QY 5281 AAACAAAAGCCGAGGAAATTTGAAGAAATATTTAACAAAAATGGAACATCATGTTGGC 5340
Db 1761 LysGlnLysThrGlnGlnAsnLeuLysLysIlePheAsnLysAsnGlyThrSerValGly 1780
QY 5341 AAAGGAAGTATGTCATACAGAAATCCCGGTAGTACTGCGGAAATTTTCTGGAAC 5400
Db 1781 LysGlnSerAspSerThrThrGlnAsnProGlnLysSerThrAlaArgLysPheThrAsn 1800
QY 5401 GAAAAATGAAGATGTGTGGAACGCAATGATATGCGGGTACAAACGCTGAGGATGAT 5460
Db 1801 GlnAsnLysGlnLysValIleThrPheAlaMetIleCysGlyTyrLysArgGlyLysAsp 1820
QY 5461 GGAATATGAGGAAATGTGCAAGAGAGTGAAGATCTAAAAAATGTGTTCTGATCT 5520
Db 1821 GlnAsnSerCysLysAsnSerAlaArgSerAspGlnAspLeuLysCysGlnSerValPro 1840
QY 5521 TCAGATGATGATTAATCTTATGGGGAATAATGGCGATGAAGGTATGCTATCATGTTCTT 5580
Db 1841 SerAspAspAspTyrProMetCysLysAsnArgAspGlnGlyThrAlaTyrGlnPheLeu 1860
QY 5581 CGATGTTTCCGCAATGGGCTGAAGATTTTTCGCAACATTAAGAAAGAAATGGAGAA 5640
Db 1861 ArgTyrPheAlaGlnTyrPheLysAspPheCysLysHisLysGlnLysGlnLeuGlnLys 1880
QY 5641 TTGTAAGGCGGCTGTAATGATTAATCTGCTGCTATTAAGATTAAGAAAGAAATGT 5700
Db 1881 LeuValGlnAlaLysAsnAspTyrThrCysGlnLysAspAsnGlnAspLysArgLysLysCys 1900
QY 5701 ACAGATGCGGTACCAATATATAAAATTTATGTAAGTGAAGCAACACTATGAATAA 5760
Db 1901 ThrAspAlaCysThrGlnTyrLysLysPheIleSerGlnTyrPheProGlnTyrGlnLys 1920
QY 5761 CAATCAAAAAATATGCTGAGAAATTAAGCAAAATATTTCCAGCATCTGTCGCAAAA 5820
Db 1921 GlnIleLysLysTyrGlnLysAsnLysAspLysIleTyrSerGlnHisProValAlaLys 1940
QY 5821 GATGCAAGAGCGCTCCGCAATATTTAGCAACAAACATTAATAAAATTTGAAAAATTA 5880
Db 1941 AspAlaGlnAspAlaArgGlnTyrLeuAspLysGlnLeuLysLysIleCysGlnAsnLys 1960
QY 5881 AGTGAAGTTGTGAATATTAAGTATGTAAGATGTGTCCACAGCATTAATGATGATGT 5940
Db 1961 SerLysAspCysGlnTyrLysCysMetLysAspAlaSerThrGlnArgLeuThrAspGly 1980
QY 5941 AATAGTCAAAATATGCCCGATCATTTAGACAGTGAACCAAAAGAGTTGAAGAAAGTGT 6000
Db 1981 AsnSerGlnAsnMetProAlaSerLeuAspAspGlnProLysGlnValGlnGlnLysCys 2000
QY 6001 AATGTCAAGTCCACGAGGTCCACACAGTGTACGAAGGAAACACCGTCACACGAGGTA 6060
Db 2001 AsnLysGlnValProArgGlnProProArgValArgArgLysThrProSerProArgVal 2020
QY 6061 TCACTGATTCAAAAGGAGAGGATGGAATAAAAGCGCAAAACAGCGCGCTTACAAA 6120
Db 2021 SerLeuIleSerLysAlaThrAlaSerLysLysGlnAlaLysThrAlaProProThrLys 2040
QY 6121 CACCGCAAAAAGTGAATAATCTAACAACAAGAAATGCGACCAACAGACCGCAGCA 6180
Db 2041 GlnProLysLysValGlnAsnLeuThrThrGlnMetArgAlaGlnThrArgThrArgArg 2060
QY 6181 GCAGCAACAAACAGGAAAGAGAACATCAACAGCAACAGCAACAGAAATCTGACGTGGC 6240
Db 2061 AlaAlaGlnGlnThrArgLysArgThrSerThrAlaThrThrGlnLysSerAspValGly 2080
QY 6241 ACAATGTAAGGCGCATTTCTTCGAATTAACAGATAGAGGAGGTGGAATAGAGGTTGT 6300
Db 2081 ThrMetValLysAlaIleLeuSerAsnLysProAspSerArgGlyLysIleGlnGlnLys 2100
QY 6301 AATCAAAAAGCTATGAGCAATATCTTAATGGGCTTATGTAGTAACTGAAGTAAAGA 6360
Db 2101 AsnProLysThrTyrGlnTyrProLysTyrPheLysLysIleValGlnLysSerLysGln 2120

Db 2841 AsnPhenValGlnLysLeuylThrPheLysSerGluCysSerValGlnSerPheSerGlu 2860
 QY 8581 TATCTTCATGAACAAGTAAGTGTGGAATTAATTAATTAAGAAATGATGCTTC 8640
 Db 2861 TyLeuHisGlnThrSerLysCysLeuasnTyLysPheasnGlnAsnAspLysSer 2880
 QY 8641 AATATACGAACATATGCTTCGAAAGAACACCAAAAGTTAAAGAAAGCTGCAGTTGT 8700
 Db 2881 AsnIleArgThrTyrlaPheGlnGlnThrProLysSerTyLysGlnLacSerCys 2900
 QY 8701 ACACATCTTCTAGAATTCATGATGATATGCTACCGATCAAAACAAAGATGATG 8760
 Db 2901 ThrLeuProSerLysasnProLeuasnPasnCysProThrAspGlnAsnLysAspLys 2920
 QY 8761 AAGGAATTCAAACTTTCACCTTCGCTGCAAGAAATGATTAATTAATTAATTCATAT 8820
 Db 2921 LysGlnLeuGlnThrPheThrPheCysSerLysAsnAspTyrlAspAsnAsnLeuAsn 2940
 QY 8821 TGGACGCGTACCTGTTCTTAATGATTCAGATGATTAACAAAGGTGATTCCTCA 8880
 Db 2941 TrpAsnAlaTyrlLeuValLeuAsnSerSerAspAsnLysGlnValLeuIleProPro 2960
 QY 8881 AGAAGAACATTTATGTACAAAGCCTATCAGTCATTAATTAATTAAGAAAGTGATTA 8940
 Db 2961 ArgArgArgHisLeuCysThrArgProIleThrAlaTyrlAsnTyrlArgLysGlyAspLys 2980
 QY 8941 GAAATTTAAAAAATCTTCTACTTCGCTTCAGTCAGCAAGCAATTTGATAGTCA 9000
 Db 2981 GluIleLeuLysLysLysLeuLeuThrSerIlePheSerGlnGlnLeuGln 3000
 QY 9001 AATATTAATGGAAGAAAGTGTGCTTGAGCAATGAATTAATTAATTAATTAATTAAT 9060
 Db 3001 LysTyrlLysSerGlnGlnGlnLeuCysPheGlnLacMetLysTyrlAsnTyrl 3020
 QY 9061 TCCATTAATTAATTAAGCACTGATGATGACACTTCATTAATTAATTAATTAATTAAT 9120
 Db 3021 SerAspIleIleLysGlnThrAspMetMetAspThrSerLeuSerGlnLysIleLysLys 3040
 QY 9121 ATATTGAAACATCAAAATGAAGCAACCGAAATCGTAAACATGAGTGGGAAATTAATA 9180
 Db 3041 IlePheGlnThrSerAsnGlnAlaThrGlnAsnArgLysThrTrpProLysAsnAsnArg 3060
 QY 9181 CGTCAGATGAGCAAGCTATGCTATGCTGATTAATTAATTAATTAATTAATTAATTAAT 9240
 Db 3061 ArgGlnIleThrPheHisAlaMetLeuCysGlyTyrlLysIleAlaThrSerLysValThrLeu 3080
 QY 9241 GATGAAGAGTGTGCAATTAACCAAGATGAAGAAATCAATCACTTCTGCTGCTTA 9300
 Db 3081 AspGlnGlyTrpCysGlnLeuProLysAspGlnGlnThrAsnGlnPheLeuArgTrpLeu 3100
 QY 9301 ATGATGAGGCAAGCAAGCATGTAAGAAAGAAACATGAAGTATTCATTAATAACA 9360
 Db 3101 IleGlnTrpAlaLysGlnIleLacLysGlnLysLysHisValSerAspSerIleLysThr 3120
 QY 9361 AATATCCGCTGTTCAACGAAGATTAATTTGAAGCTCAGATTAATTAAGCAACCTGGA 9420
 Db 3121 LysCysProArgSerAsnGlnAsnAsnPheGlnAlaSerGlnLeuIleArgLysProLys 3140
 QY 9421 TGTGATGATGATTAAGAAATATATATGCTGATTAATTAATTAATTAATTAATTAAT 9480
 Db 3141 CysGlnAsnAspIleArgLysTyrlLysSerLeuAsnIleLeuIleLysAsnThrMetGln 3160
 QY 9481 AATCTAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9540
 Db 3161 AsnLeuAsnIleLysTyrlLysGlnLeuLysAspGlnSerGlnLysAsnIleAspAsnLys 3180
 QY 9541 CCATCTGAAGAAATGTTCACTCATATATTAATTAATTAATTAATTAATTAATTAATTA 9600
 Db 3181 ProSerGlnGlnValGlnSerTyrlLysSerLysAspSerGlnCysAlaLeuGln 3200
 QY 9601 TTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9660
 Db 3201 LeuAsnAspIleAsnGlnIleValThrGlyThrLysAsnAsnGlnAsnGlnPheLys 3220

QY 9661 GAAGTACAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9720
 Db 3221 GluValLeuLysLysLeuTyrlProGlyLeuTyrlPheValGlnAspLysThrHisLysAsn 3240
 QY 9721 CATGTACTAGATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9780
 Db 3241 HisValLeuAspGlnLysAsnIleLysGlnGlnGlnThrValArgProLysAlaLeuTyrl 3260
 QY 9781 TTTCTTACACCGCATGATGATTTCTTCTATGCAAGCACTTATTCACACATCGAGTA 9840
 Db 3261 PhePheThrProHisValAspSerPheTyrlGlnAlaProLeuPheSerThrHisArgVal 3280
 QY 9841 GCACAAATGATGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9900
 Db 3281 AlaGlnTyrlAspProLysAsnAspIleLeuLysSerSerIleSerValIleValSer 3300
 QY 9901 GCGTGTAGTTTGTATGACGCTTCATTTGATGAAGAAAAATTCAAATCGTGTGACTTG 9960
 Db 3301 AlaLeuGlyLeuIleAlaLeuHisPheMetLysLysLysPheLysSerSerValAspLeu 3320
 QY 9961 TTGCGTATCTGAATATCCCGCAGAGAGATGTAATGCTGACTGTAATCCAAAAAT 10020
 Db 3321 LeuArgIleLeuAsnIleProGlnGlnLysGlyMetProThrLeuGlnSerLysAsn 3340
 QY 10021 AGGTACATACCATTAAGAGTGTCCATTAAGCAAAACATATTAATTAATTAATTAAT 10080
 Db 3341 ArgTyrlLeuProTyrlArgSerGlyProTyrlLysGlySerTyrlTyrlIleTyrlMetGlnGly 3360
 QY 10081 GATPACTAGTGAAGATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10140
 Db 3361 AspThrSerGlyAspGlnAspLysPheTyrlMetTrpAspLeuSerSerAspIleThrSer 3380
 QY 10141 TCCGAAATGAGTATGAAGAAATGGAATTAATTAATTAATTAATTAATTAATTAAT 10200
 Db 3381 SerGlnSerGlnTyrlGlnGlnLeuAspIleAsnAspIleTyrlValProGlySerProLys 3400
 QY 10201 TATAAACATTAAGAAAGTGTGCTAGAACCATCAAAAGGATTAATTAATTAATTAAT 10260
 Db 3401 TyrlSerThrLeuIleGlnValValLeuGlnProSerLysArgAspIleProSerAspAsp 3420
 QY 10261 ACACCAAGTATGATTAACACACGATGATTAATTAATTAATTAATTAATTAATTAAT 10320
 Db 3421 ThrProSerAsnAspThrProArgThrAsnArgPheIleAspAspGlnTrpAsnGlnLeu 3440
 QY 10321 AACCATGATTTGATGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10380
 Db 3441 LysHisAspPheValSerGlnTyrlLeuProAsnThrGlnProAsnAsnAsnTyrlLysSer 3460
 QY 10381 GCACATTTCCAAATGAATACAGAACTTAATTAATTAATTAATTAATTAATTAATTAAT 10440
 Db 3461 AlaAspIleProMetAsnThrGlnProAsnThrLeuTyrlSerAspAsnProGlnGlnLys 3480
 QY 10441 CTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10500
 Db 3481 ProPheIleIleSerIleHisAspArgAspLeuTyrlThrGlyLysGlnIleSerTyrlAsn 3500
 QY 10501 ATTAATATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10560
 Db 3501 IleAsnMetSerThrAsnThrAsnAsnAspIleProMetAsnAlaArgAsnAspSerTyrl 3520
 QY 10561 AGAGGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10620
 Db 3521 ArgGlyIleAspLeuIleAsnAspSerLeuValValLeuAsnLeuLeuIleTyrlMetMet 3540
 QY 10621 AAGTAT 10626
 Db 3541 LysTyrl 3542

RESULT 2
 ID 087326 PRELIMINARY; PRT; 3287 AA.
 AC 087326;

DT	01-JUN-2002	(TREMBlrel. 21	Created)
DT	01-JUN-2002	(TREMBlrel. 21,	Last sequence update)
DT	01-JUN-2002	(TREMBlrel. 21,	Last annotation update)
DE	PFEMP1 (Fragment).		
CN	TM180VAR2.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5833;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-TM180.		
RX	MEDLINE-21927235; PubMed-11930336;		
RA	Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;		
RT	"Identification of a conserved Plasmodium falciparum var gene		
RT	implicated in malaria in pregnancy."		
RL	J. Infect. Dis. 185:1207-1211(2002).		
EMBL	AF420411; CAD20867.1; -		
DR	NON_TER		
FT	3287	3287	
SO	SEQUENCE	3287 AA; 383550 MW; 58F8F866FC244536 CRC64;	

QY 1771 TTTTGAATTAATGCGTACATATTTAAGGAGTACTATTAAGTGAATGCAAACTT 1830
Db Phehegluileutrpvaltlrtylleuvalgrsphrlllelystrpansaplyleu 600
QY 1831 AAAACGTGTAATATACACACGCAATGTAATGCAATGTAATGCAATGCTTA 1890
Db LysThrcyslleasnshnrlalrhniscyslleaspilucyslyrsagancysleu 620
QY 1891 TGTATTGACAGATGGCTTAAACAAAAGAAAGAAAGAAAGAAAGAAAGAAAGTTC 1950
Db CysPheasprrgtrpvallysglnlysglnlglutrpnsnerllelyslsleuphe 640
QY 1951 ACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGTTC 2010
Db ThrlslylsasnsnvalProglinProlytrThrAsnllleasnshnleupheglugly 660
QY 2011 TATTTTAAAGTATGTAAGTAACTGACAAAGATGAAAGCAAAAGAAAGAAAGTTC 2070
Db TyrPhepelysvalmetasprlyslasprlyasnlglnlalystrpysgluleumet 680
QY 2071 GAAAAATATATAAGAAAAATGAGTTTCCAAATTTGAAAAATATAAGGACTATTA 2130
Db GluasnllelyslslysserGluPheSerAsnleuglulbasnshnrgasprlyleu 700
QY 2131 GAGAAATGCAATAGACTCTTGTAGACTTAAAGAAAGTCCCAAGTATGTAAGAC 2190
Db GluasnlalilegluleuLeuLeuaspHlsleuylslnlthralalrhllecyslysaP 720
QY 2191 AATATATACAAAGCAATGTGAACATCCATATATGCAACAAACCAACCGTGTGTTAA 2250
Db AsnasrthAsnclunlacysglnlthSerHlsAsnAlrhlthrasnProcyvallyls 740
QY 2251 CCTCGGAGGACGCAACCCCAATTAATATATAAGAAATAGCAATATCTTAAAG 2310
Db ProalrglulgllythcInProthrlsAsnlllelysglnlAlleaglntyrPheylsar 760
QY 2311 AGTGCATACGAGAAAGCAAGAAATCGTGTCTTCAATTAATGAAAGGAAAGGACACGA 2370
Db SerAlatylglnlulalargshnrgelyleuHlslyslsleuylsAlasln 780
QY 2371 GGTATATATAACGCTGGGGGTAGAAAGAAAGAAAGAAAGAAAGAAAGTATGTAATG 2430
Db GlyllelyllylsargllycllyarlgarlysaspPheylsAspAsnleucysargllemet 800
QY 2431 AATAAACAATTCATATCGTATCTGTTTCAATGGAACCAATGGAAGGCAACGCA 2490
Db IlelyshlsSerAsnshnleucllyPheSerAsnlglyProcyAspGlylysllythr 820
QY 2491 GGTGATGTATACAAACAGATTTGCTAGAACCTGAATGGAGAGTGCAGCCGGAACAC 2550
Db 821 GlyAsprglylleglInThrArghelvalyAlglyThrglnlulrvalAspProgluhls 840
QY 2551 ATGCGTAAAGATCAGAGAGATGTTATGCTGCTTACAGACGACATATATGATATCC 2610
Db MetarglylsAsprhlsngllyAspvalllemetProalrgararhlslllecystrSer 860
QY 2611 AATTGGAACAATTCACAAAGGATGATACCCACTTAATGTAATATGTTGATGATTTA 2670
Db AsnleuglulhlsleuglnlthrasprhlsProleuasnlglysllevalAlasprleu 880
QY 2671 GTTAAATTCCTTTTGGGGAGATGTTCTTCTATCAGCAAAATATGAAAGCAACAGATA 2730
Db 2671 GTTAAATTCCTTTTGGGGAGATGTTCTTCTATCAGCAAAATATGAAAGCAACAGATA 2730
Db 881 ValAsnshnsrPheleuglyAspvalleuLeuSerAlalylslyglnlalylyslle 900
QY 2731 ATAGAGATGTAAGAAAGAAAGAAATACCTAAAGGCCCCCAAGAAAGTACGACCCAAA 2790
Db 901 lleglInmetlyllysglnlylsAsnshnleuylslylProlysglnlvalthAspProly 920
QY 2791 CACAGACAACTATCTGTGAGTATAGTACGTTATGAGATATAGGATATATAT 2850
Db 921 HlsGlnlalarhllecyssarglAllearglySerPheAlasprllegllyAspille 940
QY 2851 CGAGGAAGAGATCTCTGGGAAAGAAACGGTACATGTTAAAGTCAAGACATTTGGAA 2910

Db 941 ArggllyArAspleutrpelulrghsnlglyaspmevallyslsleuclnlyhlsleugln 960
QY 2911 ACTGTTTGGTAATATACATAGTCACTCAAGGCAAGAAAGAAATGATTAATATGAT 2970
Db 961 ThrValPhegllyAsnlllelyllySerlleuylsGlylygllyAsnAsprlystrsnsp 980
QY 2971 GATGCCCAATATATTAATTTAGGGGAAATTTGGTGGAAAGCTAATAGACCCAAAGTA 3030
Db 981 AsprAlarPlystrlyleuylsleuArqgluAsprllyrPglulAlasnarqlylalyval 1000
QY 3031 TGGCAAGCCCAATTAATGATATTAATTTGAAGATTAATCGGACACCAACCAACA 3090
Db 1001 TrpGlnlaleuylsCysasprllelystrlyleuylsAsprlySerAlProglInserThr 1020
QY 3091 CAAAGTATGTTATGCGGATATAGTATACATACCAATGATGATTAATCCCAAAA 3150
Db 1021 GlnSerSerlyCysgllyllySerAsprhlsThrProleuAsprrlylleProglInly 1040
QY 3151 TTAAGATGATGACCGAATGGCGAGAAATGCTCTGCAAGTGCAGAAAAGAGATATGAT 3210
Db 1041 LeuAlgtrPmetThrGlnlulrPalaglulrPlyllyCyslylvalGlnlyllysglnlytrAsp 1060
QY 3211 AAGTTGAAGGAGAAAGTGAAGAGTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3270
Db 1061 LysleuylsGlnlyllyCyslysglnlyllyCyslylsAsprlystrlyAsnlglnlyllyCysThrly 1080
QY 3271 GAGAGTGTACAGCTGTGTACAGATGTCACAGAGCTTGAATGAATATATATATA 3330
Db 1081 GluSerGlyllyllyCysCysThrlyllyCysThrlyllyllyllyllyllyllyllyllylly 1100
QY 3331 GGATATGGAAGAAAGCAATGGAATATATATATATATATATATATATATATATATATATAT 3390
Db 1101 GlyleutrplysglnlulrPasnllleSerAsprlystrlyllyGlnleuHlsGln 1120
QY 3391 GCACAATGCTGTGTAATATAGTATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 3450
Db 1121 AlaGlnMetSerValSerAsnSerGlylleglulalSerSerAlalalyAsnHlslle 1140
QY 3451 GACAGGAATGTTATGAAATTTTGTGGAATATATATATATATATATATATATATATATAT 3510
Db 1141 AsprghasnvalileglulPheleuSerlleuylrGlnGlnAsprlyllyllyllyllyllyllylly 1160
QY 3511 AAAAGTGTACTAGTATGTAAGTGTCTGTATGCTGTACTATACACAGCTATGAAATGTT 3570
Db 1161 LysSerGlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllylly 1180
QY 3571 GGAGCATATCTCCATGATATACGAGAAATTTGATGATGTCACACAAATGAGTTTGT 3630
Db 1181 GlylalarlyleuHlsAsprhlnGlyllyAsnPhasprlysglnleuglnAsnlglnPheCys 1200
QY 3631 GATGAAAAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3690
Db 1201 AsprGlnlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllylly 1220
QY 3691 CATGATGCTGCTGTGTTGTAAGTGAATGCAACCGCAAGGATGATGATGATGATGATGAT 3750
Db 1221 HlsAsprAlalacyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllylly 1237
QY 3751 AAAAAAGGAGGAGAAAGAAAGATACGGAATGTAATAACAGTATGATATGATTAAGAA 3810
Db 1238 Glnlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllylly 1253
QY 3811 AACGATGAAAGAAAGAAAGTGAAGATGTCATCACAAGAAAGAAAGTGAATGATGATGATGAT 3870
Db 1254 Asnasnlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllylly 1272
QY 3871 GATTGCAATGC-----GGAATATATAATTTAGTGAAGACCCCTGTGTGT 3918
Db 1273 GlulrPglucylulyslyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllylly 1288
QY 3919 ATGCCCCCTGAGAGCAAAAGTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTAT 3978

QY 6097 GCGAAGAACGGCCGCTACAAACAGCGGAAAAAGTGAATAATCTAACACAGAAATG 6156
DB 2009 AAlaySargAlaProSerThrLysProLysValGluLysProArgThrGluLeu 2028
QY 6157 CGA-----GCACAACACGAAACCGAGCAGCAGCAGCAACAAACGCA----- 6198
DB 2029 GlnSerProAlaAlaGlySArgThrArgAlaAlaGlnGlnThrArgThrGlnProVal 2048
QY 6199 -----AAGCAACATCAACAGCAACAGCAACAGCAATGCGTGGCGCAAGTGA 6249
DB 2049 ArgGluAlaProGlnAlaArgThrSerThrSerThrProProAspValAlaThrMetVal 2068
QY 6250 AAGGCCATCTCTTGGATTAACCGATAGCAGGGGTGAATAGAGGGTGTGAATCCAAA 6309
DB 2069 LysSplLeuSerArgGluProAsp---LysGlyGlyLeuGluGlyLysAsnProLys 2087
QY 6310 ACATATGACAAATCCCTAAATGGGCTTGTATGTAGTGAATCTTAAGAAATGAAT 6369
DB 2088 ThrAsnGlySerThrProLysThrProLysThrProLysThrProLysThrProLys 2105
QY 6370 GCGATATGATGCTCTAGAGAGAAAAAATATGATTAATATATATATATATAT 6429
DB 2106 AlaAlaLysMetProProAlaArgLysLysLysLysLysLysLysLysLysLys 2125
QY 6430 TATGAACATGAAAAATAGCAGCAATGATATTAAGAGCTTTATTAATGTCAGCA 6489
DB 2126 AsnLysThrGluLysLysProGluLysAspLysLysLysLysLysLysLysLys 2145
QY 6490 ATGAAACATCAATTTTGGTAAATATATATATATATATATATATATATATATAT 6549
DB 2146 LLeuGluLeuThrPheThrPheLysLysLysLysLysLysLysLysLysLys 2165
QY 6550 TTGCAAAATGAGCAATTCAGATGATTAATTAAGAAATATATATATATATATAT 6609
DB 2166 LeuLysAsnGlyArgLysProAspGluProLysArgLysLysLysLysLysLys 2185
QY 6610 TATTAAGATATGTTTGGAACTGATATATATATATATATATATATATATATAT 6669
DB 2186 TyrLysAspMetPhePheGlyThrAspLysLysLysLysLysLysLysLys 2205
QY 6670 AATAGGTAAACCAATCTCTC---AATGAAATATATAGAAAAACAGATAAAAA 6723
DB 2206 AsnAsnValThrThrLysLysLysLysLysLysLysLysLysLysLysLys 2225
QY 6724 AAGATGAAGAAATTAATATATATATATATATATATATATATATATATATATAT 6783
DB 2226 LysAspGluGluLeuLeuLysAsnPheThrPheGluGluLysLysLysLysLys 2245
QY 6784 ATGATATATGATTAATCTATCATCTCACAGACGAAAAAGAAAAAATTAAGAT 6843
DB 2246 MetLysGlyLeuThrHisHisLysLysLysLysLysLysLysLysLysLys 2265
QY 6844 AATTCACATGACATGACATGACATGACATGACATGACATGACATGACATGACAT 6903
DB 2266 AsnSerGluThrThrAspLysLysLysLysLysLysLysLysLysLysLys 2285
QY 6904 CCCCATTATTTGAGATGCTCACAGATGCGCAGAAATTTTGAATAGAGAGAA 6963
DB 2286 ProGlnPheLeuArgThrPheThrGluThrPheGluGluPheLysLysLys 2305
QY 6964 CAGTGTAAATATGAGAGCGGGCTGAAGAAATATGAGTATGATGATGATGATGAT 7023
DB 2306 GlnLeuGluAsnLeuLysGluLysCysProAspLysThrCys-----SerValAspSer 2323
QY 7024 AAGACACAGAAATGCGAGCGCTGTAAACATATCAATCAATCAATCAATCAAT 7083
DB 2324 ThrLysHisLysCysGluAspSerCysLysLysLysLysLysLysLysLys 2343
QY 7084 ACTGAATATGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 7143
DB 2344 ThrGlnThrGluLysGlnArgLysPheLysAsnAspLysAspThrAspLysLys 2363
QY 7144 GATATCTCTTACTGAAGAGACATAGAGAAAGCAACATGCTCATGATATTTAAAC 7203
DB 2364 AspLysPro---LLeuLysThrLysLysLysLysLysLysLysLysLysLys 2382
QY 7204 ATCAATTAAGAAATATATGATGCAATTAAGATGCTTGAATGAAGAAACCTTCA 7263
DB 2383 ArgGlnLeuLysGluLeuLysGlyLysAsnAspCysSerCysMetGluLysProSer 2402
QY 7264 CAATCTCAAAAAC 7323
DB 2403 GlnLeuProLysThrLysSerThrSerGluProProAsnSerAsnAspMetProLys 2422
QY 7324 CTGATATGCTCCGCAACAAATTT---AACAGTGTAGTCTGCTGCAATTTCAAAAAG 7380
DB 2423 LeuAspGluValProGluGluLysGluLysLysLysLysLysLysLysLys 2436
QY 7381 GGAATATGATTCATACAAAAAATTAAGTCAATTAAGTCAATTAAGTCAATTAAG 7440
DB 2437 -----LysLeuGlnLysMetGluLysProAspProSerLysLysLys 2452
QY 7441 AAGCAGCAT 7500
DB 2453 LysSerAlaPheGlnLeuLysLysLysLysLysLysLysLysLysLysLys 2471
QY 7501 AATTTATACCTATGAGTCTCAAAAGAAAAAGAAATTAATTAAGTCAATTAAG 7557
DB 2472 LysLeuLysGlyLysAsnProAsnAspLysLysLysLysLysLysLysLys 2491
QY 7558 ---AATATCTCTGCGATCTTACAAACCTTATGACCTGATTAATATATAGAAAGA 7614
DB 2492 LLeuAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2506
QY 7615 AACCTTGTGAAGAAATAGAGAAATATGATTAAGTATATATATATATATATATAT 7674
DB 2507 AsnAlaCysLysAsn---AsnGluAsnProPheAspHisLysLysLysLysLys 2524
QY 7675 AAAAATTAAGATCTATACAGAGAAAAAAGAGATATGATGATGATGATGATGAT 7734
DB 2525 LysGluArgThrAsnArgValAspThrGluLysLysLysLysLysLysLysLys 2544
QY 7735 ATGCTTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT 7794
DB 2545 MetCysThrAsnProLysGluLysLysLysLysLysLysLysLysLysLysLys 2563
QY 7795 CTAAAAATGTTCTGTCGACATGCAAGAAATGAAGAAATGAAGAAATGAAGAAAT 7854
DB 2564 PheArgLysLeuLeuArgThrAlaAlaLysLysLysLysLysLysLysLys 2583
QY 7855 ---TCAGAGAAAGGGGTGCGCAATGCAATATGATGATGATGATGATGATGAT 7911
DB 2584 TyrThrLysAsnProLysLysArgAsnGlnLeuLysAspAlaMetLysLysSerPheAla 2603
QY 7912 GATCTGGGTGACATGATTAAGAGCAAGATATGATGATGATGATGATGATGAT 7965
DB 2604 AspLeuAlaPheLysLysLysLysLysLysLysLysLysLysLysLysLys 2621
QY 7966 CTTCCCTTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8025
DB 2622 AsnGluValGluThrLysLeuArgAlaAlaPheGluLysLysLysLysLys 2641
QY 8026 AAAAATTAAGTGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8085
DB 2642 GluAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2661
QY 8086 GCTAATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8145
DB 2662 AlaAsnArgGluLysValAlaThrLysAlaMetThrCysAspAlaProGluAspMetLeu 2681
QY 8146 TTTAAGAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 8205
DB 2682 PheLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2701
QY 8206 CATTAAGAGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 8265
DB 8206 CATTAAGAGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT

Db 2702 TyrAsnAspHisProProValaSpAspTyrIleProGlnArgLeuArgTyrMetLysGlu 2721
 QY 8266 TGGCTGCAATATATATGTAAGACAGTGAAGAAATTTGAAATAATCATGTT 8325
 Db 2722 TrpGlyValIleValCysLysIleLeuAsnGlnLysValaLysAsnAspCys 2741
 QY 8326 GATCAGCGTAAACATGTCAC---AGATGCAAGAAATGATATGTAAGAAATGTCGAA 8382
 Db 2742 GluLysCysLysLeuAsnAspLysLysCysSerAspAsnAspArgLysLysCysArg 2761
 QY 8383 CAGTGTAAACGAGATGTCAGAAATATATAATTTGTTCTTAATGGAATCTCATATC 8442
 Db 2762 SerCysLysGluLysCysLysGluTyrThrLysLeuLeuHisAsnLeuLysSerGlnPhe 2781
 QY 8443 GATATACATCAATAATAATACAAAGAAATTTGATGACACCAATATATACAAATCTCT 8502
 Db 2782 AsnIleGlnLysLysLysTyrAsnGlnLeu-----TyrThrLysIleGln 2796
 QY 8503 ACTTAT-----GATCATGTTCAAAATTTTGTACAAAG 8535
 Db 2797 AsnAsnArgArgGlyPheIleAsnAspAsnAspLysAsnValIleGluPhePheGluLys 2816
 QY 8536 TTGAAACCTTTAAAGTATGTTCTGTGAGAGCTTTTCTGATATCTTCATGAACA 8595
 Db 2817 ValLysMetIle---AsnAsnCysAsnValGlyThrProAspLysTyrLeuAspLysAla 2835
 QY 8596 AGTAAGCTGTTGAATATAATTAATGAATAAGATGATGTTCTTCATATACGAACATAT 8655
 Db 2836 IleHisCysIleHisTyrAsnPheThrGlnAsnGlyLysSer-----LysProTyr 2853
 QY 8656 GCCTTGCAAAACCAAAAGTATATAAGAGCTGTACAGCTTCTACACCTCTCTACAG 8715
 Db 2854 AlaPheAsnAsnHisProGluLysTyrLysSerHisCysSerCysTyrIle---ThrHis 2872
 QY 8716 AATCATTTGATATATGTTCTTACCGATCAACAAAGATGATGATGATGATGATGATGAT 8775
 Db 2873 HisProLeuAspLysCysProAsnTyrLysThrAsnAlaTyrCysLysThrIleArgHis 2892
 QY 8776 TTTACCTTCGCTCGAAGATGATGATGATATATATCTTGATATTTGAAACGATACCTT 8835
 Db 2893 IleAsnProCysIleThrLysAsnPheAspAsnAsnLeuIleThrThrGlyPheVal 2912
 QY 8836 GTTCTTAATAGTTCAGATGATTAACAAGGCTGATGATGATGATGATGATGATGATGAT 8895
 Db 2913 ValAspAsnLysLysAspLysAsnLysGlyValLeuValProProArgArgGlnHisLeu 2932
 QY 8896 TGTACAAGACCTATCACTGCATATATATTAAGAAAGGTATAAGAAATTTTAAAAAA 8955
 Db 2933 CysThrArgLysLeuThrGlyAsnArgTyrArgIleAsnGlnLysAspAsnLeuLysGln 2952
 QY 8956 AATCTTCTACTCTGCTTCTCACTGACGACAAATTTGAGGTCAAAAAATTAATTAATCGAA 9015
 Db 2953 AsnLeuIleAspSerAlaPheSerHisGlyMetLeuLeuGlyLysThrPheAsnAspTyr 2972
 QY 9016 GAAGAGTTGCTTGAAGGCAATGATATATGATATATATATATATATATATATATATAT 9075
 Db 2973 ThrAsnGlnLysMetGluSerMetLysTyrSerPheAlaAspTyrAlaAspIleLys 2992
 QY 9076 GGAACGTATGATGACACTTCA-----TTACTGCAAAAAATTAATAAATA 9123
 Db 2993 GlyThrAspMetIleGlyLysSerAsnIleAspAspPheAsnLysAspLeuLysGlnMet 3012
 QY 9124 TTGAAACATCAAT-----GAACCAACGAAATTCGTAAACATGTTGG 9168
 Db 3013 PheProGluAsnAsnSerGluAsnIleGlyLysThrThrIleSerArgGlnGlnTyrPhe 3032
 QY 9169 GAAATATATAGCTGATGATGACAGCTATGATATATATATATATATATATATATATAT 9225
 Db 3033 GluLysLysLysLysLysValIleThrAsnAlaMetLeuLysGlyLysGlyLysLys 3052
 QY 9226 TCAAAAGTATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 9285
 Db 3053 AsnAsnGlnGlyMetAspLysAsnTyrProCysAsnValProThrGlnAspGlyThrAspGln 3072

QY 9286 TTTCTGCTTGGTTAATGTAATGGCGCAAGCAACATGTAAGCAAAAGAAACATGTAGT 9345
 Db 3073 PheLeuArgTyrLeuIleGlnTyrPalaMetGlnLacCysLysValLysAsnHisValArg 3092
 QY 9346 GATTCATTAACCAAAATGCTCGTTCGTAACAGCAATATTTGAAAGGTGCAATTA 9405
 Db 3093 AspSerLeuLysThrLysCysArgCysSerLysGlnAspAsnPheGlnAlaSerGlnLeu 3112
 QY 9406 TTAAGACACCTGGATGTCAGAAATGATATATAGAAATATATATATATATATATATAT 9465
 Db 3113 LeuArgGlnProGluLysGlnAsnAspIleArgLysTyrIleSerLeuAsnIleLeuIle 3132
 QY 9466 AAAATATCAATGCAAAATCTTAATATATATATATATATATATATATATATATATAT 9525
 Db 3133 LysAsnSerMetGluAsnLeuAsnIleLysTyrLysGlnLeuLysAspGlnProSerGly 3152
 QY 9526 AATATACAAATTAACCATGTCAGAAAGAAATGTCAGCATATATATATATATATATAT 9585
 Db 3153 AsnIleAspAsnLysProSerGlnGluAsnValGlnInseryrIleLysSerLysTyrSer 3172
 QY 9586 CAATCGCTTGGAGTAAATGATATATATATATATATATATATATATATATATATATAT 9645
 Db 3173 GluCysAspLeuGlnLysLeuAsnAspIleAsnGlnLysLeuAspThrGlyThrLysAsnAsnGln 3192
 QY 9646 AATATCAATTCAAAGAAGTACTAAATAATATATATATATATATATATATATATATAT 9705
 Db 3193 AsnAsnGlnPheLys---ValLeuLysLysLeuTyrProGlyLeuTyrPheValGluAsp 3211
 QY 9706 GAAACACACAAAAATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9765
 Db 3212 GluThrHisLysAsnHisValLeuAsnAspLysAsnIleLysGlnGluGlnThrValArg 3231
 QY 9766 CTTAAGCACCTCTATTTCTTTACACCCCATGATGATGATGATGATGATGATGATGATGAT 9825
 Db 3232 ProLysValLeuTyrPhePheThrProHisValAspPheSerTyrGlnArgLeuPheLeu 3251
 QY 9826 TCAACACATGACGATGACATATATGATGATGATGATGATGATGATGATGATGATGAT 9885
 Db 3252 ProThrTyrArgValGluLysTyrAspProLysAsnAspIleLeuLysSerSerIleSer 3271
 QY 9886 GTTGTATATGATGCGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 9933
 Db 3272 ValValIleValSerAlaLeuGlyLeuIleAlaLeuLeuPheMetLys 3287

RESULT 3
 ID 08T325 PRELIMINARY; PRT; 1615 AA.
 AC 08T325;
 DT 01-JUN-2002 (TEMBLRel. 21, Created)
 DT 01-JUN-2002 (TEMBLRel. 21, Last sequence update)
 DE 01-JUN-2002 (TEMBLRel. 21, Last annotation update)
 GN TM284VAR3.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TM284;
 RX MEDLINE=21927235; PubMed=11930336;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babbiker H.A., Raza A.,
 RT Identification of a conserved Plasmodium falciparum var gene
 RT implicated in malaria in pregnancy.
 RL J. Infect. Dis. 185:1207-1211(2002).
 DR EMBL: AJ420412; CAD20868.1; -
 FT NON_TER 1615
 SQ SEQUENCE 1615 AA; 188639 MW; FCA896C00B86DEA6 CRC64;

Alignment Scores:
 Pred. No.: 0 Length: 1615
 Score: 7367.00 Matches: 1358
 Percent Similarity: 87.22% Conservative: 69

Best Local Similarity: 83.01% Mismatches: 169
Query Match: 37.87% Indels: 40
DB: 5 Gaps: 12

US-10-087-013-1 (1-10628) x Q81325 (1-1615)

QY 34 ATGGGAAATGCAGCATCATTTAGAGGAGATGCTAAAGCCCTATTATTAAGAAAGT 93
DB 1 MetGlyAsnThrGluSerSerLeuGluGluAlaArgSerProSerIleIleGluSer 20
QY 94 CACAAAGTGCAGAAAGTGTGACAGCTTATGCCAAAAATATAGACATCCATCAAAA 153
DB 21 GluAsnSerProArgAsnValLeuGluValArgValAlaValAsnIleArgGluAlaSerLys 40
QY 154 TATGCAAAAAGACATGCGATCGTTGAAAGGAGATTGACAAAAGCAAAATTTGGTGT 213
DB 41 AspGluAsnGluHisValAspSerLeuLysGlyAspLeuThrLysAlaGluPheArgGly 60
QY 214 GGTCTCTTACGCCAGTAAATAGCATATATTTATTCATATCCATATGTAATTTAGT 273
DB 61 GlyProSerThrProValAsnLysHisAsnLysLysLysLysLysLysLysLysLysLys 80
QY 274 CATAGGAACATCACTAATTTAGCGATATGATGATGATTTAGACATCTCTGCCATGCT 333
DB 81 HisLysGluHisLysThrAsnLeuArgTyrAspAspValAsnLeuArgHisProCysHisLysGly 100
QY 334 AGAAGAACAAACCGATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 393
DB 101 ArgGluGluAsnArgPheAspGluLysGluGluGluGluGluGluGluGluGluGluGlu 120
QY 394 TATTAAGAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 453
DB 121 TyrLysArgGluAsnAspAlaIleAlaCysAlaProProArgArgHisMetCysAsp 140
QY 454 AAAAAGCTGGAAGCTCTAATGATGATGATGATGATGATGATGATGATGATGATGAT 513
DB 141 LysAsnLeuGluAlaLeuAsnAspIleAsnThrGluAsnIleHisAspLeuGluLysAsn 160
QY 514 GTACTAGTTACAGCAAAATACGAAGTGAATGATGATGATGATGATGATGATGATGAT 573
DB 161 ValLeuValThrAlaLysLysGluGluGluGluGluGluGluGluGluGluGluGluGlu 180
QY 574 ACTTGAGACGCTTGTACGCTCTTGCAGAAATTTGGAGATATGATGATGATGATGAT 633
DB 181 ThrSerAspValCysThrAlaLeuAlaArgSerPheAlaAspIleGluLysIleLeuArg 200
QY 634 GGAATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 693
DB 201 GlyIleAspMetPheLysProAsnValHisAspLysValGluLysGlyLeuArgGluVal 220
QY 694 TTCAAGAAAATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 753
DB 221 PheLysLysIleHisAspIleMetGluGluValLysLysLysLysLysLysLysLysLys 240
QY 754 TGTGGAATTTATTAATAATAGAGAGCATGATGATGATGATGATGATGATGATGATGAT 813
DB 241 SerGlyAsnTyrTyrLysLeuArgGluAlaArgPheAsnValAsnArgAsnLysValTyr 260
QY 814 GAAGCTATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 873
DB 261 GluAlaIleThrCysGlyAlaLeuProLysSerAlaTyrPheMetGluSerGluAspAsn 280
QY 874 ACACCATTAATTTCAAAATGCTTAATGCGGCATTAACAAAGAAAGTTCCT--ACCAAT 930
DB 281 LysGluLeuPheSerTyrProLysCysGlyHisAsnAsnLysAspAspProLeuThrAsn 300
QY 931 TTAGATTTATGCTCCATATTTAGCTTGTGAGCATGAGGAGGAGAAAGTTTCCGA 990
DB 301 LeuAspTyrValProGluLysLeuArgTyrPheAspGluTyrPheGluGluLysLysLys 320
QY 991 AAAAGAAATATTAATGAAAAAGCTCAAGACCTCTGCTGATGATGATGATGATGATGAT 1050
DB 321 LysArgAsnIleLysLeuLysLysValLysAspSerCysArgAsnTyrLysGluArgGlu 340

QY 1051 TATGTAGCATATATGACATGATGATGATGATGATGATGATGATGATGATGATGAT 1110
DB 341 TyrCysSerHisAsnGlyTyrAspCysThrThrIleThrPargLysGlyIleLeuHis 360
QY 1111 TTGATTAATTAAGTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1170
DB 361 LeuAspAsnLysCysThrAspCysSerThrLysCysLysValPheGluValTyrPheGlu 380
QY 1171 AATCAACAAGAAAGCTTTAAACAAAAGAAAGAAATGAAAGAAATACATCATAT 1230
DB 381 AsnGluGluGluAlaPheLysLysGluLysGluLysGluLysGluLysGluLysGlu 400
QY 1231 TTATGCAAGCATATACAAATTTGCTCAATATATTAATGATGATGATGATGATGAT 1290
DB 401 LeuSerAsnAspAlaLysPheValAsnAsnIleAsnSerGluTyrTyrLysGluPheHis 420
QY 1291 GAAAACTTAAGAAACCGCATATGCAACTAATGACACTTTTAAATTTACTAAATGAA 1350
DB 421 AspGluLeuArgAspLysAsnTyrLysAsnLeuAspThrPheLeuAsnLeuLeuAsnGlu 440
QY 1351 GGAAGTATTTGAAAGGAGATTTACACGAGAAAGAAATTAATTAATTAATTAATTA 1410
DB 441 GlyLysTyrCysLysGluLysLeuLysGlyGluAsnAspIleAsnPheThrAsnSerSer 460
QY 1411 GATGACAAAGGAGATTTTATGCTTCAGAAATTTGCCAAGTGTGCCGACGCGGGCTC 1470
DB 461 AspAspLysGlyThrPheTyrArgSerGluTyrCysGluValCysProAspGlyVal 480
QY 1471 AATGTGATGATTAATTAATACACACAAATGATGATGATGATGATGATGATGATGAT 1530
DB 481 LysArgAspGlyThrLysLysLysLysLysLysLysLysLysLysLysLysLysLys 500
QY 1531 GAAAGCTATTAACCTCCATGCGGCTGTAAGCTACTAATATACAGCTGCTTATAGTGT 1590
DB 501 GluAspTyrLysProProTyrPheLysProThrAsnIleThrValLeuTyrSerGly 520
QY 1591 AATGACAAAGGTGATTAATACACAAATTTAGAAAATTTTGTACAGCTCAATATAC 1650
DB 521 AsnAspGluGlyAspIleThrGluLysLeuGluAspPheCysLysSerSerThrAsnTyr 540
QY 1651 AAAAGTAAATAAATCAAAAAATGGGAATGCTTTTAAAGGAGAAATTAATTAATGAT 1710
DB 541 LysAspLysAsnAsnGluLysTyrPheLysGlyTyrLysAspGluAsnIleAsnArgCys 560
QY 1711 AAACGTGAACAAATATCTGAATCAATATGATTAATCTTAAGATATATCATTTATAT 1770
DB 561 LysLeuGluGluAsnThrGluLysAsnAspAsnProLysIleThrSerPheHisAsn 580
QY 1771 TTTTGTGAATTAATGCTTACATATTTATTAAGGATCTATTAAGTGAATGACAACTT 1830
DB 581 PhePheGluLeuTyrValThrTyrLeuLeuArgAspThrIleLysTyrAsnAspLysLeu 600
QY 1831 AAAAGCTGATTAATTAATCAACCAAGCATGATGATGATGATGATGATGATGATGAT 1890
DB 601 LysThrCysIleAsnAsnThrThrThrHisCysIleAspGluCysAsnArgAsnCysLeu 620
QY 1891 TGTTTGACAGATGGTTAAACAAAAGAAAGAAAGATGATGATGATGATGATGATGAT 1950
DB 621 CysPheAspArgTyrPheLysGluLysGluGluGluGluGluGluGluGluGluGlu 640
QY 1951 ACAAAAAATAATATACAGATCGTATTAATGATTAATTAATTAATTAATTAATTAAT 2010
DB 641 ThrLysLysAsnAsnValProGluProTyrTyrThrAsnIleAsnAsnLeuPheGluGly 660
QY 2011 TATTTTATTAAGTATGATGATTAACCTGACAAAGATGAAGCAAAATGGAACACTTAT 2070
DB 661 TyrPhePheLysValMetAspLysLeuAspLysAsnGluAlaLysTyrPheGluLeuMet 680
QY 2071 GAAATATTAATAAAGAAAAAATGAGTTTCCATTTGGAATTAATTAATTAATTAATTA 2130
DB 681 GluAsnIleLysLysLysLysAsnGluPheSerAsnLeuLysAsnAsnArgAspTyrLeu 700

QY	2131	GAGATGCAATAGAAACCTCTGTTAAGATTCACCTTAAAGAAACCGCAGATATGTAAGAAC	2190
Db	701	GIUSNAlAlIeGIUeUeUeUaSPHLSleUySGUthPAlAtHrIleCysLysASP	720
QY	2191	AATATATCAAAACGAAAGCATGTGAAACATCCCATATATGCAACAAACCCCTGTGTTAAA	2250
Db	721	ASnaSnthrASnglUAlaYrGIuThrSerHISaSnAlaThrLYaSPnProTYaLlys	740
QY	2251	CTCTGTGGAGGACGCAACCCACTAAAAATATTAAGAAATATAGCAACATACTTTAAAGC	2310
Db	741	ProATGcGIyGIYthrGIuPnProThrLYaSnIleUySGUleAlAGIntYrPheUeYarG	760
QY	2311	AGTCGATATCGAGGAAAGCAACAATTCGTGGCTTCATAAATTAAAGAAAGGACACAGAA	2370
Db	761	SerAlaYrGIuGlnUAlaAaYnaAaAGGILeUeHISLysLeUySGILySaLHnISglU	780
QY	2371	GGTATATATTAACGTGGGGGTAGAGAAAGGACTTCAGAGCAACATTTATGTAGAAATATG	2430
Db	781	GIYleYrLYrYsaHrGGLyGIYArGIYLSaSPHelySaSPHnleUcYsaHrGILleUeC	800
QY	2431	ATTAACACTTCATAACGTAAATCTTGCTTTTTCAAATGACACTGTGATGGCAAGGCACA	2490
Db	801	ILeUySHISerASnaAaYnaAaLeUcGIYrPheSerASngLYrProCYaSPaRGIYLySGILYthr	820
QY	2491	GGTATGTGTATACAAACAATTTGTTCGTAAGAAATCGAATGGGAAGTGATCCCGAACC	2550
Db	821	GIYaSPcGIYleGIuThrAaYnaPheValIGLYthrGIuThrGILyAlaSPnProGlnHIS	840
QY	2551	ATCGCTAAAGATCAAGAAAGATGTTATATATGCTCTAGAAAGCAACATATATGATATCC	2610
Db	841	MeArGIySaSPHISglUaSPeValIleMeCProProArGIaYrGIaHISIleCYsThrSer	860
QY	2611	AATTTGGAACATTTTACAAACGATGATCAACCCACTTAATGTAAATATTTGTTGATGATTTA	2670
Db	861	ASnleUcGlnHISleUcGIntHrASaSPHISrProleUaSnGLYASnIleValAlaSPeU	880
QY	2671	GTTATATATTCCTTTTGGGGGAGATGTTCTTCrTACAGCAAAATATGAAGCAACAAGATA	2730
Db	881	ValSaSnASerPheleUcGILyAaSPeValIleUeUeSerAlaLYsTYrGILyAlaLYaLYsIle	900
QY	2731	ATAGCAATGTTTAAAGAAAGATTAACCTTAAGGGCCCCAAAGAAAGTAACTGACCCAAA	2790
Db	901	ILeGIuMeCYrLYrSGILySaSaSnAlaLeUySGILYrProLYaSGILyAlaThrASpProLYs	920
QY	2791	CACGAGACACTATCTGTGAGGCTATACGTACAGTTTGTGCAGATATAGATGATATTAAT	2850
Db	921	HISGlnAlaThrIleCYsaYrAlaIleArGIYSerPheAlaASpIleGLYASpIleIle	940
QY	2851	CGAGGAGACATCTCTGGGAAAGAAACGGTACATGTGTAACCTGCAGACACTTTGGAA	2910
Db	941	ArGIYArGIaSPeUeThrPGLyAaSaNGILYASpMeValLYleUcGlnGLYHISleUcGln	960
QY	2911	ACTGTTTTGGTATATATACATATAGTCAACCAAGGCAAGAAAGATTAATATATATGAT	2970
Db	961	ThYrAlaPheGLYASnIleIYrLYrYSerleUySGILYySGILYASnaSPlySaThraSP	980
QY	2971	GATCCCCCAATATTTTAAATTTGAGGAAAAATGGTGGGAAGCTATATAGACCAAAAGTA	3030
Db	981	ASpIaSPnProLYrLYrLeUySleuArGIuASpTrPTrGILyAlaASnaArGIAlaLYsAl	1000
QY	3031	TGGGACCGCATGAATGTGATATATTAATATTTGAAGAGATTAATCGGACACCATCAACA	3090
Db	1001	TrPTrGILyAlaMeLYsCYsaSPrIleUySYrLYrLeUySaSPlySerArGProGlnSerThr	1020
QY	3091	CAAAAGTATTTATCCGATATATAGTACATACCACTTGAGATGATTAATCCCAAAAA	3150
Db	1021	GIuSerSerLYrCYsGLYrYrSerSaSPaSnThrProleUaSaSPlyrIleProGlnLYs	1040
QY	3151	TTAAGATGATGACCGAATGGCGAGATGTACTGCAGAGTGCACAAAAAGAGATGAT	3210
Db	1041	LeuArGIrPheCThrGIuThrPAlaGIuThrPlyrCYySaLYaGlnLYaLYeGILyArSP	1060
QY	3211	AAGTTGAAGGAAAGTCTTAAGGAGTGAAGATTAAGGATTAATGTGTCAAGCCTGTACGAA	3270

D	1061	lysleuylsglulyscylslysglucylsasp	lytrsncllyelmclgyls	thrlys	1080																			
Q	3271	GAGAGTGGTACAGGTTCTACGAGTGCACAGAGCTTGTAAGCATATATATGATTAATA			3330																			
D	1081	gluserglythrlyscyls	thrlyscyls	thraspalacysasnclutyrasnsp	lletle 11000																			
Q	3331	GGATTATGGAAGAACAATGGATATATATATCATGATAAATACAAAGAAATTACATGACAA			3390																			
D	1101	glyleutrp	lysglulntrp	asnlelle	lesertrasp	lytylsgluldeu	hls	gln 1120																
Q	3391	GCACAATGCTGTACTAATATAGTGGATATGAGCTTCCAGTACGCGCAAAATCATATA				3450																		
D	1121	Ala	glnmet	serVal	Ser	asn	Ser	glylle	gla	la	ser	Val	ala	lys	asn	hls	tle 1140							
Q	3451	GACAGATGTTATTTGAATTTTGTGCGATTTATACCAAAATAGTGTGCAAAATGAT				3510																		
D	1141	Asp	arg	asn	Val	lle	gln	phe	leu	ser	gln	leu	tyr	gln	hls	asp	lytyl	ys	ser	asn 1160				
Q	3511	AAAAGTGTACTAGTATGATGAAGAGTGTCTCATTTGGTACTACACACAGTATGAATGT				3570																		
D	1161	lys	ser	gly	thr	ser	asp	gln	ser	thr	Val	lle	gln	tyr	asn	thr	arg	tyr	gln	asn	Val 1180			
Q	3571	GGAGCATTCCTCCATGATACAGAAATTTGATGATTCGACGTCACAAATGCACTTTGT				3630																		
D	1181	gly	ala	tyr	leu	hls	asp	thr	gln	lys	asn	phe	asp	asp	gln	leu	hls	asn	leu	hls	phe	lys 1200		
Q	3631	GATGAAAAATGATGATGATTAAGGATACGAAAAATATATGCTTTAGAGATTAACACAGAC				3690																		
D	1201	asp	gln	lys	ser	asp	gln	lys	asp	lysglul	tyr	ala	phe	arg	asp	thr	pro	asn	sp 1220					
Q	3691	CATGATGGTGGCTGTGTTAAAGTGATCGAACCACAGGAGGTACAGTAAAAACG				3750																		
D	1221	His	asp	Val	ala	cyls	ala	cyls	ala	asn	gln	arg	asp	lys	-----	lle	gln	trp	pro	lys 1237				
Q	3751	AAAAAAGCGGAGAAAGATACCGATGCTAAAAACGTAAATGATTAATTAAGAA				3810																		
D	1238	gln	lys	lys	lys	gln	-----	ala	cyls	ala	lle	val	lle	gln	leu	leu	leu	lys	gln 1253					
Q	3811	AACGATGAAAGAAACAGTAAAGATTCATCACCACAAAAAGATATGATATGATATCC				3870																		
D	1254	asn	asn	gln	lyr	thr	thr	lys	val	gln	gln	cyls	asn	val	lle	tyr	thr	gln	gln	lys	leu	---tyr	trp	Pro 1272
Q	3871	GATTGGCAATGC-----GGAATATTAATTTAGTGAAGACCCGTGTGT				3918																		
D	1273	gln	trp	gln	cyls	gln	lys	lys	lle	gln	asn	lle	hls	-----	lys	gln	lyl	acys 1288						
Q	3919	ATGCCCTACAAACAAAGTATGCGTACATTTCTGGCAATGATATGATATAAA				3978																		
D	1289	Met	trp	pro	ala	arg	trp	lys	leu	cyls	Val	hls	phe	leu	lys	gln	-----	leu 1304						
Q	3979	AAATTCACAATCAAAATTAATTTAAAGAGCTTCATCAATCTGACACACACAAACA				4038																		
D	1305	lys	val	lle	gln	trp	asp	gln	lys	leu	arg	gln	la	phe	lle	gln	cyls	ser	la	la	leu	thr 1324		
Q	4039	TTCTTCTCATGATATTAATATAAAGTAGAGATGTGAAGAAATGCACTGCATAAAGA				4098																		
D	1325	Phe	leu	leu	trp	lys	tyr	lys	gln	u	asn	asn	gln	yl	gln	u	asp	re	u	asn	gln 1344			
Q	4099	TTAAAAAGAGCAAAATTCCTCCCGCATTTTGAATCCATGTCTACACATTTGGAAT				4158																		
D	1345	leu	gln	ser	gly	lle	lle	pro	asp	asp	phe	lys	arg	gln	met	phe	tyr	thr	phe	lle	asp 1364			
Q	4159	TATAGAGTTTATTTATTTGAAACGATATATCAAAAAGTCATGCTGAGGAGAGTAACTA				4218																		
D	1365	tyr	arg	asp	leu	cyls	leu	gln	lys	asp	lle	gln	ser	asp	val	gln	-----	asn	val 1381					
Q	4219	AAAGACAATAGATCTCTTTCAAAATGTGACCAAAATCTCTATATGAAAAACA				4278																		
D	1382	asn	gln	lys	lle	asn	ser	ala	leu	lys	lle	gln	yl	gln	ser	asp	leu	asn	sp	-----lys 1399				
Q	4279	CGCCAAAGATGTGGACAGAACATAGTCATGAGATATGGCAAGCTATGCTATGCCTACTA				4338																		


```

Db 1400 ArgLysAsnTrpTrpAsnGlyIleLysGluAspValTrpLysGlyMetValCysGlyLeu 1419
QY 4339 -----GTAATAATGGGCGCAAAAAGATGAT 4365
Db 1420 SerHisAspValSerGlyAsnHisLysGluIleValArgLysLysIleMetGluAspPro 1439
QY 4366 TTATCCGAAACATACGGTTTCACACACGTAATTTAGTACAAA-----AGCACCCT 4419
Db 1440 GluAsnAsnLysTrpGlnTrpAsnSerValLysPheThrAspGlnProSerGlyThrLys 1459
QY 4420 TTGGAGCAATTTCCCAACACACCGAGTTTTCAGATGGTACCGAATGGTACGACGAC 4479
Db 1460 LeuSerGlnPheLeuLysValProGlnPheLeuArgTrpLeuThrGlnTrpTrpAsp 1479
QY 4480 TATTCGTATACACGACAAAATTTTGAAGCATGTCACGAAAATGTAGTCAAAATGAC 4539
Db 1480 TyrCysHisThrArgLysGlnLysGlyValGluSerThrCysLysSerAsnAsp 1499
QY 4540 CAATTGAGTGTGATCAGATGTATAGAAATGGAGACACACGTTAATATG--- 4596
Db 1500 GlnLeuLysCysAspThrGlnLysAsnLysLysCysGluAspTyrGlnLysTyrMetLys 1519
QY 4597 AAAAAAAGAGTGTGATTCACAAAGATTAATATTACAGGATGACGCGCAAAAA 4656
Db 1520 LysLysLysGlnTrpLysGlnAspLysTyrTyrLysAspGlnArgAspLysLys 1539
QY 4657 AATTCGATGACACACACATGTGTGTAATGTACAGACTATACGGAAGCAATGCACAA 4716
Db 1540 ArgPheAsnArgGlnHisIleGlnValMetValLysAspTyrThrGlyThrAsnAlaThr 1559
QY 4717 GATTACTGACAGAAATTTACTGCTAGTGTGATGATGATGATGATGATGATGATGATGAT 4776
Db 1560 AspTyrLeuLysArgLysPheThrAlaSerCysGlyAspLysProGlyLysSerVal 1579
QY 4777 GTACAAAGAAATATACAAATTTAGAAAAACAGGCTTACTATGATGCCACAAACATGT 4836
Db 1580 ValGlnArgAsnIleGlnMetLeuGlnLysGlnAlaTyrTyrAspLysAspLysHisCys 1599
QY 4837 GGGTCACAAAATTTATGCAAAAATGACGCAAAATATACATGATTTGCG 4884
Db 1600 GlyCysThrLysPheIleGlnAsnAspAspLysTyrThrAsnIleSer 1615

RESULT 4
ID 015870 PRELIMINARY; PRF: 2706 AA.
AC 015870:
DT 01-JAN-1998 (TREMblrel. 05 Created)
RT 01-JAN-1998 (TREMblrel. 05, last sequence update)
PI 01-DEC-2001 (TREMblrel. 19, last annotation update)
PI EMBL (Fragment).
R239+VAR1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5633;
RM NCBI_TaxID=5633;
RP SEQUENCE FROM N.A.
RC STRAIN=IT 4/25/5;
RX MEDLINE=97373957; PubMed=9230440;
RZ Rowe J.A., Moulds J.M., Newbold C.I., Miller L.H.;
  "Plasmodium falciparum rosetting is mediated by PLEMP1 and requires
  complement receptor 1."
RL Nature 388:292-295(1997).
DR EMBL; Y13402; CAAT3831.1; JOINED.
DR EMBL; Y13403; CAAT3831.1; JOINED.
DR InterPro; IPR002086; Aldehyde_dehydrt.
DR Pfam; PF03011; PEMP; 2.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
FT MON_TER 2706 2706
SQ SEQUENCE 2706 AA; 308162 MW; 1C3D55AD5317D68 CRC64;

```

```

Score: 4515.50 Matches: 1191
Percent Similarity: 41.69% Conservative: 372
Best Local Similarity: 31.77% Mismatches: 871
Query Match: 23.21% Indels: 1315
DB: 5 Gaps: 123

US-10-087-013-1 (1-10628) x 015870 (1-2706)
QY 100 AGTCGCAAGAAATGTTTGGACGTTTTCGCAAAATATACACATTCATCAAAA----- 153
Db 14 SerAlaThrAspValLeuGlnLysIleAlaThrGlyIleTyrAsnGlnLysGlnLys 33
QY 154 ---TATGCAAAAGACATGTCGATTCGTTGAAGGGGATTTACGAAACGACAAATTT--- 207
Db 34 ValTyrProTyrGlu-----AsnGluLeuLysGlyIleLeuSerAsnAlaIlePheVal 51
QY 208 -----CGTGGTGGCTCTTACGCGCAATTAATAG 237
Db 52 AspGlnLeuArgLysGlnLeuAsnIleGlnSerProGlyProSerAsp----- 67
QY 238 CATTAATTAATTAATCAATATCCATGTAATTAATTAATTAATTAATTAATTAATTAATTA 297
Db 68 -----SerCysSerLeuAspHisLysPheHisThrAsnIleAsn 80
QY 298 TATGATGATGTAATTTGACACATCCTTGGCATGGTAGAACAACCAATTTGATGTA 357
Db 81 ThrGlnTyrThrGlnGlyArgLysProCysTyrGlnArgAsnGlnLysArgPheSerAsn 100
QY 358 GATGCAAGATTCGATGTCGGA---AATAAATACGTAATTAATTAATTAATTAATTAATGATGCT 414
Db 101 GlnGlyGlnLysLysCysGlySerAspLysIleArgAspTyrGlyIleLysSerAlaGly 120
QY 415 ATAGCCTGTGCGCCACCTGAGAGACACATATGTTGTAATAAAGTTGACGCTTAAT 474
Db 121 GlyAlaCysAlaProPheArgThrGlnAsnLeuCysAspArgAsnLeuGlnTyrLeuIle 140
QY 475 GATTAATAATACCCAAATATTCATGATTTATGGAATGACTACTAGTACGACAAATATC 534
Db 141 AsnLysAsnThrAsnThrThrHisAspLeuGlnLysValLeuValThrAlaLysTyr 160
QY 535 GAAGTGAATCAATGTTAATTAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 582
Db 161 GlnGlyAspSerIleValAsnAsnHisProAspLysAsnSerSerGlyAsnLysSerSer 180
QY 583 GCTTGTACTGCTTTCGACGAAGTTTTCACATATAGTGTATGTTTGAAGCAATGATGAT 642
Db 181 IleCysThrAlaLeuAlaArgSerPheAlaAspIleGlyAspIleValArgGlyArgAsp 200
QY 643 ATGTTTAACCAATATGTCATGACAAAGTACAAACGGGCTCCGACAGAGTTTTCAGAAA 702
Db 201 MetPheLysProAsnAspAlaAspLysValGlnLysGlyLeuGlnValAlaPheGlyLys 220
QY 703 ATACATGATGATGAATGAAGATGAAGTAAATAAATGATTAAC---AATCCGATGATGTCGGA 759
Db 221 IleTyrAsnSerLeuProSerProAlaGlnLysHisTyrAlaHisAspAspGlySerGly 240
QY 760 AATTAATTAATTAATTAAGAAAGATGTCGATGATGATGATGATGATGATGATGATGATGAT 819
Db 241 AsnTyrTyrLysLeuArgGlnAspTrpTrpAlaIleAsnArgLysGlnValAlaTrpLysAla 260
QY 820 ATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 879
Db 261 IleThrCysArgAlaProAsnGlnAlaAspPheThrArgAsnIleSerGlyAsnMetLys 280
QY 880 TTATTT---TCAATTCCTAATATGCGGCGCATTAACAAAGAAAGTTCCTACCAATTTGAT 936
Db 281 AlaPheThrSerGlnGlyTyrCysGlyHisSerGlnThrAsnValProThrAsnLeuAsp 300
QY 937 TATGCTCCCTCAATTAATTAATGCTTGGTTCGACGAATGGGAGAGAGTTCGCCAAAAAGA 996
Db 301 TyrValProGlnPheLeuArgTrpPheAspGlnTrpAlaGlnGlnLysPheCysArgGlyLeuArg 320
QY 997 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1056

```

Alignment Scores:

Pred. No.: 1,2e-219

Length: 2706


```
OY 3157 TGGATACCGAATGGCAGAAATGCTACTGCAGAGTGCAGAAAAAGAGTATGATTAAGTTG 3216
Db 722 -----CysAsp11eLeuProLysSerValAspAsp 732
OY 3217 AAGGAGAAGTGAAGAGGTGAAGTAAAGATTAATGTCACAGGCTGTACGAAAGAGCT 3276
Db 733 GUGGluAspAspGluGluVal ----- 739
OY 3277 GGTACAGGTTGTACGAAGTGCACAGAGCTTGTAAATGAATATATATATATAGATTAA 3336
Db 739 ----- 739
OY 3337 TGGAAAGCAATGATGATATATATATATATATATATATATATATATATATATATATAT 3396
Db 740 -----AspGluGluLysGlu ----- 744
OY 3397 ATGTCTGTAGTAATAGTGTATGTAAGCTTCCAGTACAGTCCCAAAATCATATAGCAGG 3456
Db 745 -----GluSerSerGluThrThrLysArgAsnIle ----- 754
OY 3457 AATGTATTGAATTTTGTGCGAATTTATACAAATAATGTCGCAAAAGTATATAAGT 3516
Db 755 -----SerGluLysGly 758
OY 3517 GGTACTAGTGAAGAAAGTCTGCTCATGTTGCTACTACACACAGTGAATAAGTTGGAGCA 3576
Db 759 GylThrLysSerAlaSerCysValLysGlyAlaCysAlaIleValLysGlyVal -----Leu 777
OY 3577 TATCTCCAGTATACAGAAATTTTGTATGTCAGTGCACAAATAGTTTGTGTATGA 3636
Db 778 GlnGlnLysSerAsnLysSerIleAspAsnCysAsnAlaLysAsnArg ----- 793
OY 3637 AAAAGTATGATGAAGATACGAAAAAATATGCTTTAGATTAACACAGACCATGAT 3696
Db 793 ----- 793
OY 3697 GGTGCGTGTGTTGAAGTGTGATGCAACGACAGAGGTACAGTAAACGAAAAA 3756
Db 793 ----- 793
OY 3757 AAAGCGAAGAAAAAGATACGGAATGTAAACAGTGAATGATATATTAAGAAAAAGAT 3816
Db 793 ----- 793
OY 3817 GCAAGAAACAAGTACAGATGTCATCCAAAAAAGATAGTAATGATATCCGATTCG 3876
Db 794 -----LysLysAsn -----GluTrp 798
OY 3877 CAATCGGAATATATAATTTAGTGAA -----GACCCCTCGTGTGTATGCCCCCTTGAAAG 3933
Db 799 GlnCysAspLysAsnThrPheValAspGluAsnGluGlyValCysMetProProArgArg 818
OY 3934 CAAAAGTATGCGTACATTTTGGCAATGATTAATGAATTAATAAATTAACATCAACA 3993
Db 819 LysSerIleCysIleHisAsnLeuThrLeuGluGluGlnThrLys -----AsnLys 835
OY 3994 GTTAATTTAAAAAGAGCTTTTCATCAATTCGACGACGACGACGACGACGACGACGAC 4053
Db 836 TyrGlnLeuArgGluAlaPheIleLysCysAlaAlaLysGluThrAsnLeuLeuTrpAsp 855
OY 4054 TATTAATAAGTAAGATGGTGAAGAAATGAACTCGAATTAAGAAATTAAGAAAGCAAA 4113
Db 856 LysTrpLysAsnAspLysAsnGluAlaGluLeu -----LeuLysLysGlyLys 872
OY 4114 ATTCCTCCCGATTTTGAATCCATGTTCTACACATTTGGAGATTATAGATTTTGA 4173
Db 873 IleProGluAspPheMetArgIleMetPheThrThrPheGluAspPheArgAspPheCys 892
OY 4174 TTGGAACAGTATATCAAAAAGCTCATGTCGACGACGACGACGACGACGACGACGACGAC 4233
Db 893 LeuGluAsnAspMetGlyLys -----AspValAspLysValLysLysAsnIleAsn 909
OY 4234 TCTCTTTCAAAAATGTCGACCAAAAATCTCCTAATGAAAAAACACCGCAAGATGTCG 4293
Db 910 LysValAlaPheAsnAsnSerSerLysArgLysPheLysLysIleAspProGluAsnTrp 929
OY 4294 ACAGAACATATGTCAGATATATGGAAGCTATGCTATGTCACATAGTAAAAATGGGCA 4353
Db 930 AsnGluAsnGlyProGlnIleThrPheAsnGlyMetLeuCysAlaLeuIleHisAlaThr 949
OY 4354 AAA -----AAAGATGATTTACCGAAAAACTAGCGTTTACAAACCTGCAAA 4398
Db 950 LysAspSerIleLysAsnLysAsp -----AsnTrpLysTrpGluLysValThr 965
OY 4399 TTTAGTACAAA -----AGCACACTTGGAGAAATTTGCCAAAGACCC 4443
Db 966 IleLeuAlaLysArgAspLysSerAsnGlyMetThrLeuSerGluPheAlaLysLysPro 985
OY 4444 CAGTTTTCATGCTGCTTAACGGAATGTAACGACGACTATGCTATACACGAAAAATAT 4503
Db 986 LysPheLeuArgTrpPheValGluTrpLysAspArgLysGluLysGluLysTrp 1005
OY 4504 TTGAAGGATGTCAGGAAAAAATGTAAGTCAATGAC -----CAATTGAAGTGTATACA 4557
Db 1006 LeuThrGluValAlaSerThrCysLysSerIleAspGlyGluLysLeuLysCysAspArg 1025
OY 4558 GAATGTAATTAAGAAATGCGAGGACTACGTTAATATATATGAAAAAATAAAT -----GAGTGG 4614
Db 1026 GlyCysAsnAsnLysCysAspLysTrpLysLysTrpMetArgLysLysLysLysLysLys 1045
OY 4615 ATTCACAAAGTAAATATTTACAAAGATGACCGCAAAAAAAGTTCGATAGCAACAC 4674
Db 1046 AsnLeuGlnAspLysLysTrpLysAspLysArgGluAsnLysGlyIleAspLysGlyPro 1065
OY 4675 ATGTGTAATGTTTACAGACTATCTGTAACGACGACGACGACGACGACGACGACGACGAC 4734
Db 1066 IleGlyIleIleValLysAspTrpValLeuAlaAsnAlaLysGluThrLeuLysLysLys 1085
OY 4735 TTTACTGCTAGTTGTGCTGAATAGCTTGAAAGTGC -----TCGTGCTGA 4779
Db 1086 PheThrAlaSerCysValThrSerSerGlyLysAlaGlnAsnSerAlaThrGluGluVal 1105
OY 4780 CAAAGAAATATACATTTTGAAGAAACAGCTTCTATGATGCCGACAAACATTTGTGG 4839
Db 1106 LysLysAsnIleGluLeuLeuSerGluGluGlnTrpLysAspAlaAspGluTrpCysGly 1125
OY 4840 TCCACAAATTTATGAAATGACGACAAATATATACATTTTGCAGTAAGTAAGTCG 4899
Db 1126 CysThrLysPheIle -----HisAspLysTrpSerLysIleSerGlyArgSerAsnCys 1144
OY 4900 AAAGATTAATGAAGGACGACAAACACAGTCTATTAAGTGCACAAACMAAGTCTTAAT 4959
Db 1145 CysGlyLeuAsnSerAspAlaLysLysAsnLysIleLysTrpArgAsnSerAspGluLys 1164
OY 4960 AACTCAATTAACCTTGAAGAA -----TTGACTGAAGTGTGCTTTTCTTCCTTCGTCGA 5013
Db 1165 AspTrpAlaPheLeuLysLysArgAsnLeuSerGlyAspValPhePheProSerArgArg 1184
OY 5014 CTACGATATGTTTTCATGACATTTGATGCGCAATTAATACAGATCCAGAAATTAAGATGAA 5073
Db 1185 LeuArgIleCysPheHisAlaLeuAspGluLysAsnTrpThrAspProGluValLysAspGlu 1204
OY 5074 AATGGGTCGCAAAAAGTTATGAGAGTGGCGCACGACGACGACGACGACGACGACGACGAC 5133
Db 1205 AsnGlyLeuArgLysArgLeuMetGluValAlaIleAlaThrGluGlyTrpAsnLeuSerGln 1224
OY 5134 TACTACAAAGAAAAAAGAAAAAGAAAAATTAATAAAGCTGCGATGGCGCAAAATATCT 5193
Db 1225 TyrTrpLysGluLysLysGluLysGluThrGluAlaThrGluGluAlaHisLysTrpSer 1244
OY 5194 TATAGGTCGCGCTTGTAGTCTATGAAATATAGTTTATGATTTTAAAGATATATAT 5253
Db 1245 TyrGluValGlnProCysSerAlaMetLysTrpSerPheTrpAspLeuArgAspIleIle 1264
OY 5254 CTAGTATGATTAATTTGGAAGATGAAAAACAAAGACCGAGGAAAAATTTGAAGAAAAATA 5313
```


QY 7012 ---AGTAAT-----GACGGTAAGACACAGAAATGT 7038
Db 1971 CysSerAsnIleAspValaAsnLysAspLysIlePheHalaAspLeuGluGlyProLysCys 1990
QY 7039 GCAGAGCGCTGTGTATCAAAATTTATTAAGAGAGGAAAGAACTGAAATATGAAGA 7098
Db 1991 AlAlaLysProCysSerSerIYrAqLysTrIleLysGlyLysLysTrIleGlyGluLys 2010
QY 7099 CAAGAGCAAAAGTTCAAAAAGATTAAGATGGCAAAAAGTATAG-----GATTAT 7149
Db 2011 GluGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2030
QY 7150 CCTTCTACCTAAGAGCATAGAGAGGCAACATGTCTCATGAAATTTTAAACATGAAGAA 7209
Db 2031 AsnGlyPheArgValLysLeuGlyThrCysThrThrAlaGlyAspPheLeu----- 2047
QY 7210 TTTAAAGAAATTAATGATTAAGATTTGTTCTGTATGACAAAACCTTTCACAACAT 7269
Db 2048 ---GlnThrLeuLysAsnGlyProCys----- 2055
QY 7270 CCAAAAACACACACACATCACAATCCGATCGTAATGATATGCGCAGAAATCGTGAT 7329
Db 2056 ---LysSerGluAsnGlyLysAspHisGlu----- 2065
QY 7330 TATGTCCTGAAGATTTAAACAGTGTGAGTGTCTCGAAGCTTCAAAAAAGGATCTATG 7389
Db 2066 ---AspGluLysAsnPheSerGln---ProAspValThrPheArgProAla--- 2080
QY 7390 ATTCAATACAAAAAAATTAACCTTAAATACCTATGATGATGTAGAGAAAGCAGCA 7449
Db 2081 ---ThrAsnGlyThrProCysProLysPheLysValAsnGly----- 2093
QY 7450 TATATTTATCTAAAGACAGAAATAATATGATATTACCTTGAGAGAAATTTTATA 7509
Db 2093 --- 2093
QY 7510 CCTATTAGCTACAGAGAAAGAAAGAAATAGTTGGACTAATATATATCTCTGC 7569
Db 2094 ---LysAsnGlyAsnGlyAlaAspThrAsnGlyLysCys 2106
QY 7570 GATCCTACAAAACCTTATGCACCTGATAATATATAGAGAAAGAAACCTTGTGAAAT 7629
Db 2107 AsnGlyLysThrPro----- 2111
QY 7630 AGAGAGAAATCGTTTAAGTAGATTTATGATGGAATGTACAAAATTCAAAATTC 7689
Db 2111 --- 2111
QY 7690 TATCAGAGAGAAAAAGAGTATGTACTCCAGAGAGAAATATGTCTTAAGAAAT 7749
Db 2111 --- 2111
QY 7750 TTAGATGAATTAATTAAGAACTTAAGATAGTATATTTCTCTAAAATGCTTGT 7809
Db 2112 IleAspAlaGlnAsnPheGluMetGlyGlnThrAlaLysGluPheValMetLeuVal 2131
QY 7810 CGAACTGCAGCAAAATGAAGATAGACATATATAAAAAATTCAAGAGACGGCTGC 7869
Db 2132 SerAspLysSerThrAsnGlyPheGluVal-----AsnAspLeuAsnGluLys 2147
QY 7870 GCATGAATCCAAATGTGATATCATGAAATATAGTTTCGTGATCGTGAGTGCATAGTT 7929
Db 2148 Ala----- 2148
QY 7930 AGAGAGACAGATATGTACGAATGTGTGTTACTTACTCCCGTAGAAATTAATATATAT 7989
Db 2149 ---GlyAlaAsp----- 2151
QY 7990 AAGGTTTGAATCATATATAGAGAAATGAGAGAAATTAATAAGGTAGAAATTAATATAC 8049
Db 2152 ---IlePheGlnGlyIle----- 2156

QY 8050 AACGATGACAAAAGTTTGCTTCTGCTGGTGGATGTATATAGAAAGATATTTGAAA 8109
Db 2157 --- 2157
QY 8110 GCATAT-----ACGTGCAAGACACAGAA-----GATGCAAAA 8142
Db 2163 CysArgAsnValCysGlyLeuAspValCysLys---ProGluLysValAsnAspGlnLys 2181
QY 8143 CTTTAAAGAAAGAGAAATGATGATTTGAACGATACATTAATACAAATTAAGTGT 8202
Db 2182 ValAsnGlyLys---GluAsnAspGly----- 2189
QY 8203 GGACATTAAGAGAGATCCACTGTGATGATATATATCTCAACGGTTTGATGATGACT 8262
Db 2190 ---ThyTrIleIleGlnIleArgAlaLeuArg 2200
QY 8263 GAATGCTCGAATATATTTGTAAGACAGTGAAGAAATGAAAAATTTAAATATCA 8322
Db 2201 ArgTrpValGluAsnPhe-----LeuGluAspTrpLysLysIleLysHisLys 2216
QY 8323 TGTGATCACTGTAAAAACATCTGCAGATGCAAGATGATATATGATGAAATTAAGTGTGA 8382
Db 2217 IleSerHisCysThrAsnSerThrGluGluLys-----LysSerThrCysAsp 2232
QY 8383 CAGGTAAACGAGATGTCACAGATATATATATATATATATATATATATATATATAT 8442
Db 2233 ---CysGlyLysLysCysLys-----CysValGlyGlnTrpIleLysLeuLys 2247
QY 8443 GATATATCAATTAATTAATACAAAGATTTGTATGAAACACAAATATATATATATAT 8502
Db 2248 LysGluGluTrpLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2258
QY 8503 ACTTATGATCATGTCAAAATTTGTACAAAAGTTGAAAACCTTTAAAGTGAATGTTCT 8562
Db 2259 ---ValGlyLysLysSerGlu----- 2265
QY 8563 GTTGAGAGCTTTTCGAATATCTCATGAAACAGTAAGTGTGAAATTAATATATAT 8622
Db 2266 ---AspAspPheSerAspAsnLeuAsnSerPheLeuGlnThrLeuIleThrGlnIlePro 2284
QY 8623 GAAATGATGTTCTTCAATATACAAATATATGATGTTTGAAGAAACCAAAAAGTAT 8682
Db 2285 ValAlaAspValGlnGlyAsnValIleLysLeuSerAsnPheAspThrPro----- 2301
QY 8683 AAAGAGCTTGCAAGTTGTACACTACCTTCAAGATTCATGATTAATGTCTACCGAT 8742
Db 2302 ---CysGlyCysSerAlaAspAlaAsnSer----- 2310
QY 8743 CAAAACAAAGATGATGTAAGGAATTAACAACCTTTACTCTGCTGAGAAATGATATAT 8802
Db 2311 GlnLysLysAspGly-----AsnGluAsnAspAla 2320
QY 8803 GATATATATCTGATTAATTTGAAGACATACCTGTCTTAATATGATGATTAACAAA 8862
Db 2321 IleAspCysMetIleAsn-----ArgLeuGlnGlnLysAlaLys 2333
QY 8863 GGTGATTTGATTCCTCCAGAGAGACATTTATGTAACAGACCTTACACTGATATATAT 8922
Db 2334 AsnGlySerLysGlnGlnProSerGlySerLysGlnCysThrProThrThrLeu--- 2352
QY 8923 TATGAAAAAGGTGATTAAGAAATTTTAAAAAAACTTCTTACTTCTGCTTCAAGTCA 8982
Db 2352 --- 2352
QY 8983 GGACAAATTTAGTCAAAAATATATATCGAGAGAGAGTGTGCTTGAAGCAATGAAA 9042
Db 2353 ---GluAspAspGlnThrPheAsp----- 2359
QY 9043 TATAGTATGAGATTTATTCGATATATATTAAGAGAACTGATATGATGACACTTCATTA 9102
Db 2360 ---AspAspIle----- 2362
QY 9103 TCTGAAAAAATTAATAAATTTTGAAGACATCAAAATGACAGAACGAAATCGTAAACA 9162

```

Db 2363 -----GluThrAspSer----- 2366
QY 9163 TGGTGGGAAATATAGACGTGAGATATGCGACCCATGTTATGATATATAATTCCT 9222
Db 2367 -----ProValAlaHisProGlnIleCysGluGlyValIleGln 2379
QY 9223 ACTTCAAAAGTAACTATAGATGAAGCATGGTGTCAATTCACCAAGATGAAGAACTAAT 9282
Db 2380 ThrGlnThrGluThrValIleGluGluGlyValIleProProIleGluGlyVal 2399
QY 9283 CAGTTTCTGTTGGTAAATGAATGGGCAAGACATGTAAAGAAAGAAACATGTA 9342
Db 2400 Asp-----GluLysLysLysPro 2405
QY 9343 AGTGATTCATTAAACCAAAATGCTCTGTCAACGAGATATATTTGAAGCGCTCAGAA 9402
Db 2406 GluAspPro----- 2408
QY 9403 TTATTAAGACACCTGGATGTCAAGATGATTAAGAAATATATTTACCTGAATATATGG 9462
Db 2408 ----- 2408
QY 9463 ATAAAAATATCAATGCAAAATCTAATATATAAGCAATTAAGATCAATCTTCA 9522
Db 2409 -----AlaGluGluAspGly 2413
QY 9523 GGTAAATATAGACATATA-----CATCTGAGAAATATCTCATATATATA 9570
Db 2414 GlyAlaIleGlyProSerGlyProIleGlyProGlnIleProSerProIleHis 2431
QY 9571 AAATCAAAAGATTTCTCAATGCCCTTGGAGTTAAATGATTAATGAATATGTTACAGCA 9630
Db 2431 ----- 2431
QY 9631 ACAAAAATATATGAATATGAATTCAGAGTACTAAAAATATATATCTCGTTTA 9690
Db 2431 ----- 2431
QY 9691 TATTTGTGAGATGAAGACACAAATATCATGTACTAGTGGAAATATATAAGAAAGAA 9750
Db 2432 -----AspAspThrGluGluAsnHisVal 2439
QY 9751 GAGCAACACGTGCTCTAAAGCAGCTATTTCTTACACCCCATGTAGATCTTTTGTAT 9810
Db 2440 -----ThrProGluGluAspProPro 2447
QY 9811 CAAGCACCCTTATTCACACACATGAGTAGACAA----- 9846
Db 2448 ProIlePro-----AspThrArgProValProSerProIleProProIleAlaAspGln 2465
QY 9847 ---TATGATCCTAATAAATGATATATATAAGTATCTCTGTGTTATTTGATTCGCG 9903
Db 2466 ProPheAspPro-----ThrIleLeuGlnThrThrIleProPheGlyValAlaLeuAla 2483
QY 9904 TTAGGTTGATAGCGCTTCATTCATGAGAAAAATTCAAATGCTGTGGAGATTGTTG 9963
Db 2484 LeuLeuSerIleAlaPhePhePheMetLysLysThrLysHisProValAspLeuPhe 2503
QY 9964 CGTATATGATATCCCGCAGAGAGATGATGGAATGCTGCTGGAATCCAAATATAGG 10023
Db 2504 SerValIleAsnThrProLysGlyAspLysAspIleProIleLeuLysSerLysAsnArg 2533
QY 10024 TACATACCATATAGAAAGTGTCCATATAAAGCAAAATATATATATATGAGAGAT 10083
Db 2524 TyrIleProLysArgSerGlySerLysGlyLysThrLysIleLysMetGluGlyAsp 2543
QY 10084 ACTGTGAGATGAAGTAAATATATGCGACTATCTCTGATATATCTTCAATCC 10143
Db 2544 Ser-----AspGluAspLysTyrAlaPheIleSerAspThrAspIleThrSerSer 2561
QY 10144 GAAGTAGATGTAAGAAATGATATATATATATATGACAGGATGCTAAATAT 10203

```

```

Db 2562 GluSerIleLysGluGluLeuAspIleAsnAspIleTyrValProGlySerProLysTyr 2581
QY 10204 AAACATTGATAGAGTAGTACTAGAACCATCAAAAAG-----GATATACCA---AGT 10254
Db 2582 LysThrIleuIleGluValIleuGluProSerLysSerAsnGlnIleProHisSer 2601
QY 10255 GATGATACACCACTAATGAT---ACACGACGTACGATATGATTTATGATGATGAATG 10311
Db 2602 AlIGlyLysLysProLeuAspAspMetValProThrThrAsnThrPheThrAspGluGlyTyr 2621
QY 10312 AATGATGAAGATGATTTGTTATCTCATATATATACCAATACAGAACCAAAATATAT--- 10368
Db 2622 AsnIleuLysHisAspPheIleSerGlnThrIleGlnSerArgLeuProMetAspVal 2641
QY 10369 ---AATTACAAA---AGTCGATATATTCATGATAGATACAGAACTAATATATATCT 10422
Db 2642 ProGlnIleAspValSerThrGluLeuProMetAsnIleValGlyAsnValLeu---Asp 2660
QY 10423 GATATCTCGAAGAAAACCTTTTATATATCTATTCATGATAGGATTTATATCTGGG 10482
Db 2661 AspLysIleAsnGluLysProPheIleThrSerIleHisAspArgAsnLeuLysAsnGly 2680
QY 10483 AAAGAAATTAATGTTATATATATATATATAGTAGTAACTAATATATATGATATTCATGAT 10542
Db 2681 GluIleLysSerLysAsnIleAsnMet-----ValAsnAsnAspIleProMetSer 2697
QY 10543 GCTAGAAATGATTTCTTATAGAGGTATA 10569
Db 2698 AspLysAsnGlyThrTyrThrGlyIle 2706

RESULT 5
Q26032
ID Q26032 PRELIMINARY; PRT; 3006 AA.
AC Q26032;
DE 01-NOV-1996 (TEMBUREL. 01, Created)
DT 01-NOV-1996 (TEMBUREL. 01, last sequence update)
DT 01-DEC-2001 (TEMBUREL. 19, last annotation update)
DE Variant-specific surface protein.
GN VAR-3.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MMDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
RA Peterson D.S., Ravetch J.A., Wellens T.E.;
RT "The large diverse gene family var encodes proteins involved in
RT cytoadherence and antigenic variation of Plasmodium falciparum-
RT infected erythrocytes.";
RL Cell 82:89-100(1995).
DR EMBL; I40609; AAA75397.1;
DR InterPro; IPR004258; PfEMP.
DR Pfam; PF03011; PfEMP; 2.
SQ SEQUENCE 3006 AA; 343769 MW; 8FD5FF475F943C74 CRC64;

Alignment Scores:
Pred. No.: 8.63e-207 Length: 3006
Score: 4260.00 Matches: 1143
Percent Similarity: 43.84% Conservative: 483
Best Local Similarity: 30.82% Mismatches: 1005
Query Match: 21.90% Indels: 1079
Gaps: 134

US-10-087-013-1 (1-10628) x Q26032 (1-3006)
QY 91 AGTACAAAGTCAAGCAATGTTTGGACGTTATGCCAAATATATAGCAT---CCA 147
Db 8 ThrAsnLysThrAlaLysGluValIleuAspIleLysGluValGlnGluValAla 27
QY 148 TCAAAATATGCAAAAGAAATGCTGATTCGTTGAAGGGATTTGACAAAGCAGATTT 207

```


Db 28 ThrGluAspAlaLeuThrTyrArgAsnAspLeuGlnGlyAsnLeuSerGlnAlaIleAspHe 47
QY 208 CGTGGTGGCTCTTACCGCAGTAATAGCATATTTATTTATCCATATCCATGTAAT 267
Db 48 HisGly-----ValProIleAspValIleAsn-----ProCysAsp 59
QY 268 TTAGATCTAAGGAACATCTAATTTACGGTATGATGTAATTTGAGACATCTTGC 327
Db 60 LeuAsnTyrGluIleHisThrAsnVal---ValIleGlyArgGlyLysGluAsnProCys 78
QY 328 CATGTGAGAAACAAACCGATTTGATGAAGATGAAGAACTGATGGCAATATAATA 387
Db 79 ArgGlyArgGluGlnIleValArgPheSerAspValLeuSerGlnGlyAlaIleAsnThr 98
QY 388 CGTAAATTAATAAGAAAAATGATGCTATAGCCTGTGCGCCACCTAAGACAGCATATG 447
Db 99 IleLysAspSerValIleHisSerValIleValAcylalProPheArgGlyLeuHisVal 118
QY 448 TGTGATAAAACTGGAAGCTCTAAATGATTAATATCCAAAATATTCATGATTTATG 507
Db 119 CysAspTyrAsnLeuGlnLeuIleLysThrAspGluAsnThrSerThrHisAspIleLeu 138
QY 508 GGAATGTACTAGTACAGCAAAATACGAAGTCAATCATGTTGTAATATCAATCCACAT 567
Db 139 ValAspValIleuLeuThrAlaLysHisGlnGlyLysSerLeuValIleHisTyrLysGlu 158
QY 568 -----AAGAACTTCAGAGCTTGTACTGCTCTTCACAGAACTTTTGACAT 615
Db 159 TyrIleLysLysAsnArgAsnArgPheAsnIleCysThrValIleuAlaArgSerPheAlaAsp 178
QY 616 ATAGTGTATTTGTAAAGAGATGATATGTT-----AAGCAAAAT 657
Db 179 IleGlyAspIleValAlaArgGlyLysAspLeuTyrLysGlnTyrAspArgGlnIleLysGlu 198
QY 658 GTCCATGACAAATATAGAAACGGCTCCGAGAGCTTTCAAGAAATACATGATGCAAT 717
Db 199 LysArgLysGlnLeuGlnLysAsnLeuLysPhePheGlnLysIleHisAspAspVal 218
QY 718 -----GAAGATGAAGTAATAATGATTTCAAT 744
Db 219 MetLysThrSerLysArgThrAsnGlnLysLysSerAlaGlnAlaGlnLysArgTyrAsn 238
QY 745 CCGATGATCTGGAATATTTATTAATTAAGAGAAACATGTTGTAATGTAATGAAT 804
Db 239 ---AspAlaThrGlyAsnTyrLysLysLeuArgGluAspTyrPheAsnAlaAsnArgAsp 257
QY 805 AAGATGTGGAAGCTATACATGATGATGATATATTAATTCGATATTTATGCAATCA 864
Db 258 GlnValTyrLysAlaIleThrCysAspAlaAlaAspAsnAspGluTyrPheGlnLysSer 277
QY 865 GAAAGTAATACACATATTTCAATCTAATGCGGCATTAACAAAGAAAGGTTCT 924
Db 278 SerAspGlyLeuTyrValPheSerAsnGlnLysGlyArgAsnGlnIleLysValPro 297
QY 925 ACCATTTGATTAATGCTCCCTCAATATTTAGCTTGGTCGAGAAAGGAGAGATTT 984
Db 298 ThrAsnLeuAspTyrValProGlnHisLeuArgTyrPheAspGluTyrAlaGlnLysPhe 317
QY 985 TCCCGAAAAAGAAATATTAATTTGAAGAAAGCTGCTCTGCGTAATGCAAAAGAA 1044
Db 318 CysArgLysArgAsnIleThrLeuLysSerAlaLysGlnLysCysGlyGlyLysLysAsn 1044
QY 1045 CCCTATATTTAGTCATTAATGACATGATGTAACACATATTTGGAAGAAAGATTT 1104
Db 338 ThrLysTyrCysSerLeuAsnGlnLysArgPheValPheGlnLysLysAspSer 357
QY 1105 TTGCATTTGGAATTAATAGTACTGATGCTGCGATAATGCAAGTTTGAAGTTTG 1164
Db 358 CysSerSerAspGlyAsnCysThrAlaCysSerAsnIleCysIleValTyrAspAlaTyr 377
QY 1165 TTAGGGAATCAACAAGACATTTTAAACAAAGAAAGAAATATGAAAGAAATACAA 1224
Db 378 LeuArgAsnGlnArgAsnGlnPheGlnLysGlnLysIleLysTyrThrLysGlnIleGln 397

QY 1225 TCATATTTATGACAGATTAACAATTTGCAATATTAATTAATGATATTAATACAA 1284
Db 398 LysTyrLysSerSerSerAspLysSerAsnSerAsnIleSerAsnLysTyrTyrAsnGln 417
QY 1285 TTTTGAAGAAAACTTAAGAAACGCAATATGCAATATGATGACATTTTAAATTTACTA 1344
Db 418 PheTyrGlnAsnPheGlnLysLysGlnTyrGlnThrLeuGlnAsnPheLeuLysLeu 437
QY 1345 AATGAAGAAAGATTTGTAAGAGAGATTAACAGAGAAAGAGATTAATCTTTACTAAC 1404
Db 438 AsnLysGlyMetTyrCysGlnGlnLysIleGlnGlnGlnIleValIleAspPhe----- 455
QY 1405 AGTGTGATGACAAAGGATATTTATGCTGAGAAATATGCAATATTTGCCAAGTGTCCGACTGC 1464
Db 456 AsnLysAspGlnAspMetValPheHisArgSerGlnTyrCysGlnProCysProAsp 475
QY 1465 GGGGTCAATGTGATGATTAATTAATTAACACACAAATGATGATTAATGCGTAACGT--- 1521
Db 476 ValValGlnCysLysGlnLysGlnTyrGlnAspLysLysAsnAspLysCysArgSer 495
QY 1522 -----GTAAATTAAGACATTAACCTCCATGGGCTGTGAAGCT 1563
Db 496 LysIleIleLysLysIleLeuGlnSerGlnIle-----Pro 507
QY 1564 ACTAATATGATGCTGCTTATATGATGAAGAGATATTAACAAATTAAGAA 1623
Db 508 ThrGluIleHisValIleuAsnSerAspAspLysGlnLysPheIleThrLysLysLeuGln 527
QY 1624 AATTTTGTACAGCTCAATTAATTAACAAGATTAATAATTAATTAATGCAATGCTAT 1683
Db 528 ValPheCysSerSerThrThrAsnTyrGlnGlnArgAsnValGlnIleTyrLysCysTyr 547
QY 1684 GTATGAGATGAATAATTAATTAATGATTAAGTGAACAAATATGCAATATGAT 1743
Db 548 AsnLysAsnSerAspTyrAsnAsnCysGlnMetAsnIleSerSerTyrLysAspSerThr 567
QY 1744 AATCTTAAGATTAATATCATATTTTGAATTTGGTTACATATTTATTTAGG 1803
Db 568 AspAlaAsnValMetLeuSerValGlnCysPheHisSerThrAlaLysAsnLeuIle 587
QY 1804 GATATATTAATGGAATGACAAACTTAATACTGTATTAATTAATTAACA---ACCAGCAT 1860
Db 588 AspThrIleLysTyrGlnHisGlnLeuLysAsnGlnLysIleAsnAsnThrAsnValThrTyr 607
QY 1861 TGTATTTGATGATGATTAACAGAAATGCTTATGTTGACAGATGCGTTAAACAAAGAA 1920
Db 608 CysGlnSerLysCysIleLysAsnCysGlnCysTyrGlnLysTyrIleLysArgLysGln 627
QY 1921 GAAAGATGGAATGATTAAGAAACGTGTCACAAAAAAGAAATATTAACGCAATCGAT 1980
Db 628 HisGlnTyrGlnLysValLysAsnValPheGlnLysAsnAsnArgMetSerTyrIleTyr 647
QY 1981 TATGTAATATTAATTAATCTTTTGAAGTTATTTTAAAGTATGATTAACCTTGAC 2040
Db 648 TyrAsnAsnLeuSerAlaValPheAspSerPheLeuPheGlnValMetPheAlaLeuAsp 667
QY 2041 AAACATGAA---GCAAAATGAAAGAACTTATGCAAAATTAATAAGAAAGAAAGAAAG 2097
Db 668 GlnAspLysGlnLysTyrAspGlnPheThrGlnAspLeuLys---LysLysPheGln 686
QY 2098 TTTTCCAAATTTGAAATTAATAGGACATTTAGCAATATGCAATCTTTGTTAGAT 2157
Db 687 ProSerLysThrAsnThrProThrGlyLysSerGlnAspAlaIleGlnPheLeuAsp 706
QY 2158 CACTTAAGAAACCTGCCAGATATTAAGACATTAATTAACAAAGAGATGTAACAA 2217
Db 707 HisLeuLysAspAsnAlaLeuThrCysArgAspAsnAsnSerAsnGlnLysCysAspVal 726
QY 2218 TCCCATTAATCAACAACAAACCGTGTGTTAACTCTGTGAGAGACGCAACCATTA 2277
Db 727 SerLysLysValLysThrAsnProCysGlnLysAsnProSerAlaSerAsnLeuVal 746

QY 2278 AATATAAAGAAATAGCACATACTTTAAAGAGCTGCATACAGACAGCAACCAATCT 2337
Db 747 ArgValIysrGleuAlaGluMetGlnArgIyrlaArgIyIsleuGluLysArg 766
QY 2338 GGTCTTCAT---AAATGAAAGAAAGAAAGCAGACAGAGTATATATAACGTGGGGAGG 2394
Db 767 G1yG1yG1uIleAsnLeuLysG1yAspAlaThrLysG1yThrArgG1nG1yG1yPro 786
QY 2395 AGAAAGACTTCAGACAAATTTATGTAGATAATGATAAACATCTCT-----AAT 2445
Db 787 AlaAspIyIleLys---AsnValIySerIleAsnGlnAsnHisThrAsnValGlnAsn 805
QY 2446 CGTATCTGTGCTTTTCA---AATGACCATGTGATGGCAA---GCCACAGCTGATGCT 2499
Db 806 AsnAsnArgIalArgIyrlArgIyngIyProCysThrIyIyAspG1ySerAsnG1yG1y 825
QY 2500 ATACAAACAGATTTGCTGAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 2559
Db 826 Val-----ArgMetLysIleG1yThrProIyLys---ProG1yArgG1nIleG1nMet 842
QY 2560 GATCAGCAGAGATTTATTTATGCTCCTAGAAAGACATATATGATATGATATGATGAA 2619
Db 843 SerAlaG1uAspIleIyMetProIyArgIyGlnHisMetCysThrSerAsnLeuG1u 862
QY 2620 CATTTAAACAGATGATCCACTT---AATGTAATATGTGTATGATTTAGTTAAT 2676
Db 863 TyrIleuG1nThrLysAspG1yProIleuLysG1nIyAsp-----G1yLysLeuValAsn 880
QY 2677 AATTCCTTTTGGGGAGTGTCTCTATCAGAAATATATAGCAAAACAGATTAATGCA 2736
Db 881 AsnSerIheLeuG1yAspValIleuLeuSerIalLysMetAspAlaG1yLysIleIleG1u 900
QY 2737 ATGTATAAGAAAGAAATTAACCTTAAGGGCCCAAGAAAGTAACTGACCCCAACACAG 2796
Db 901 LeuTyrLysLysG1nAsnAsn-----LysSerAsnLeuThrAspProG1uAspAsn 917
QY 2797 ACAACATCTGCTGATATAGTATAGTTGCAATTTAGCTATAGCTATATATATGCGAGA 2856
Db 918 GluSerIalCysArgAlaLeuAlaArgIySerIheAlaAspLeuG1yAspIleIleArgIy 937
QY 2857 AGAGATCTCTGGGAAAGAAAGCGTACATGATGTAAGCTGCAGAGCAATTTGGAATCTGT 2916
Db 938 ArgAspLeuThrAspIyAsnSerAspAlaLysArgIleuG1nThrAsnLeuLysG1uIle 957
QY 2917 TTGCTATATATACATAAGTACTCAAAAGCAAGAAAGTAAATATATATGATGATGCC 2976
Db 958 PheThrLysIleLysG1nG1uLeuProG1uAspIleLysLysLysTyrAspLysAspG1y 977
QY 2977 CCCAAATATTTAAATTTGAGGAGAAATTTGCTGGAACCTATATAGCCCAAGTATGGGAA 3036
Db 978 ThrAspHisLysLeuLeuAlaG1uAspIyIyProG1uAlaAsnArgHisGlnValIyIyArg 997
QY 3037 GCCATGAATGTGATATAATATATTGAAGATTAATCGGGACCAATCAACACAAGT 3096
Db 998 AlaMetLysCysAlaIleG1uAsnAspLysAspMetLys----- 1010
QY 3097 AGTATTGCGGATTAAGTATACATACCACTTGATGATTTATATCCACAAAATTAAGA 3156
Db 1011 -----CysAsnG1yIleProIleG1uAspIyIleProG1nArgIleuArg 1025
QY 3157 TGGATGACCGAATGGGAGATGCTAGTGCAGAGTGCAGAAAAGAGTATGATTAAGTGT 3216
Db 1026 ThrPheThrIleuIy 1045
QY 3217 AAGGAGAAGTGAAGAGTGAAGATGAAGATGAAGTGAAGTGAAGTGAAGTGAAGTGA 3276
Db 1046 ValAlaAspCysLysSerCysLysG1yLys-----AlaLysSerCysThrGlnLysAsp 1063
QY 3277 GGTACAGTTGTACAGAGTGCAGAGCTTTGTATGATTAATGATTAATGATTAATGATTA 3336
Db 1064 G1y---AspCysThrLysCysLysAlaIalCysAspAsnIyAsnLysLysIleLysPro 1082
QY 3337 TCGAAAGAACATGGAATTAATTAATCAATTAACAAAGATTAACATGAACAGCAAA 3396

Db 1083 TrpIleuG1uIleIyIyIyIyLysAsnLysTyrIalIleuIyLysLysAlaLeu 1102
QY 3397 ATGCTCTGTATATAGTATGATTAATGAAGCTTCCAGATCCCAAAATCATATAGACAGG 3456
Db 1103 AspSerValAsnG1yLysG1uIyLysLysLysLysThrLysSerAspAlaLysAspGln 1122
QY 3457 AATGTTATGCAATTTTGTGCGAATTTATACCAACAAAT---GGTGGCAAAAGTAAATA 3513
Db 1123 GlnValValHisPheLeuAlaG1uIleuIleArgLysSerIyG1yG1yLysG1yLys 1142
QY 3514 AGTGTACTAGTATGAAGTGCCTGATGTGATCTATACCAACCACTGATGAATAATGTGGA 3573
Db 1143 AsnValLysThrThrValSerProThrThrProAsnThrLeuIySerSerAlaAla 1162
QY 3574 GCATATCTC---CATATACAGCAAAATTTGATGATGTGCTAGTCCACAAATGAGTTTGT 3630
Db 1163 G1yTyrIleHisIySg1uLeuG1yArgThrValG1yLysAsnThrGlnLysG1uPheCys 1182
QY 3631 GATGAAAGAGTGAAGTAAAGATACGAAATAATATGCTTTAGAGATTAACACAGAC 3690
Db 1183 TyrSerLys-----AsnG1yLysTyrIalPheLysAspProIyLysG1y 1197
QY 3691 CATGATCTGCTGTGCTGTTAAAGT---CGATCGAAACCGACAGCTACAGATATAA 3747
Db 1198 TyrG1uG1uIalCysLysCysAsnAspArgAsnProLysPro-----Gln 1212
QY 3748 ACGAAAAAAGCAAGCAAGAAAGATACGCAATGTAAACAGATGATATACTTAA 3807
Db 1213 ProAlaProLysLysG1uAspG1uAspAla---CysAspValValLysProLeuLeuLys 1231
QY 3808 GAAAACGATGAAAGAAACAGTGAAGATTCATCA-----AAAAAATAGTAAAT 3861
Db 1232 AspLysG1yG1uThrAspAlaIleAspG1yCysAsnGlnLysTyrIalG1yLysAsp 1251
QY 3862 GGAATGCCATGCGCAATGCGGAATATATAATTTAGTGAAGACCTCGTGTGATG 3921
Db 1252 LysTyrProG1yIyIyAspCysAsnSerGlnIleHisThrHisAsnG1yAlaCysMet 1271
QY 3922 CCCCTGAAGACAAAAGTATGCTATCTTTGCAATGTAATGAATGAATGAATGAATGA 3981
Db 1272 ProIyArgIyGlnLysLysCysValSerG1yLeuThrLysThrAspArgIleLysAla 1291
QY 3982 TTACATACACAGTTAATTTAAAGAAAGCTTTCATCAATCTGCAGACAGCAAAATTC 4041
Db 1292 IleG1u-----TyrIleArgThrGlnPheIleLysSerIalAlaIleG1uIyHis 1308
QY 4042 TTCTCATGTATTTATTAAGTAAGTAAAGTGAAGAAATGACATCGATTAAGATTA 4101
Db 1309 PheAlaIyIyAspArgIyLysG1uAspAsnG1yG1uAla-----GluAlaG1uLeu 1325
QY 4102 AAGAGAGCAAAATTTCCCTCCGCAATTTTGAAGATCCATCTTCACATTTGGAGATTA 4161
Db 1326 LysAsnG1yLysIleProG1uG1yPheLysArgIleuIyIyIyIyIyIyIyIyIyIyIy 1345
QY 4162 AGAGATTTTATTTGCAACAGATATATCAAAAGCTGATGCGAGGAGTAACTTAA 4221
Db 1346 ArgAspIlePhePheG1yArgAspIleSerThr---HisAlaTyrIleSerG1yValSer 1364
QY 4222 GAGCAAAATAGATCTCTCTTTC---AAAAATGCTGACCAAAA-----TCT 4283
Db 1365 ProLysValIleThrIleLeuG1uLysG1uAsnAspAlaLysTyrIalAlaLysGlnAsn 1384
QY 4264 CTAATGAAAAAACAAGCAAGATGCTGACACAGACATAGTCAATGATTAATGGAAGCT 4323
Db 1385 SerAsnAsnG1uLeuLeuAspAspIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 1404
QY 4324 ATGCTATGTCACATAGT---AAATTTGG-----GCAAAAAAAGTATTTATCCGAA 4374
Db 1405 MetLeuCysAlaLeuThrHisLysIleSerAspG1uIyLysLysG1uIleLysAsn 1424
QY 4375 AACTACGGTTACCAACAGCTCAAAATTTAGTGACAAA---AGCACACCTTTGAGAGATTT 4431

Db 1425 LysTyrSerTyrLysLysLeuAsnGluSerProLysGlySerAsnLysValGluAspPhe 1444
QY 4432 GCCAAGCAGCCCGGATTTTACGATGGCTAACGGAATGGTACGCGCATATGCTATAC 4441
Db 1445 AlAlLysLysProGlnPheLeuAlaGlyPheLeuIleGluTyrGlyAspGluPheCysAlaGln 1464
QY 4492 CGACAAAATTTTGAAGGATGTCAGAGAAAATGT-----MAGTCMAATGAC--- 4539
Db 1465 ArgGluGluLysGluAlaLysValLysValSerCysSerAspAlaLysAspTyrAspGly 1484
QY 4540 -----CAATGAGTGTGACAGACAGATGTATATAGAAAATCCGAGCATCTTAATAT 4593
Db 1485 CysLysAsnThrLysSerAsnAlaSerCysValSerAlaCysLysValTyrGluAspTyr 1504
QY 4594 ATGAAAAAAGAGAGAGTGCATCCACAGATTAATATTACAGAGATGAACGCGACAAA 4653
Db 1505 IleThrLysLysLysVal-----GluTyrThrLys-----GlnLys 1516
QY 4654 AAAAGATTGATGACACACATTTGTGTATGTTTACAGACTTACTGA----- 4704
Db 1517 GlyLysPheAspAlaGluLysIle-----ThrAspLysGluGlyTyrGluGly 1532
QY 4705 -----ACGATGCAAGATTTACTTGAACAGAAAATTTACTGCTAGTTGCTGAT 4755
Db 1533 PheSerThrLysAspAlaSerGluTyrLeuLysLys-----Cys----- 1546
QY 4756 AAGCCTGGAGTGCCTCTGTGTACAAAAGAAATATACAAATTTAGAAAACAGGCTTAC 4815
Db 1546 ----- 1546
QY 4816 TATGATCCGACAAACATTTGGTGCACAAAATTTATTAAGATGACGCAATATACT 4875
Db 1547 -----LeuAspAspThrCysAsnCysMetGlnLysAsnThrGluTyrTyr 1564
QY 4876 AACATTTCCAGTAAGATTAAGTCAAAAGATTAAGAGGCAACACAGGCTGAT 4935
Db 1565 AsnThr----- 1566
QY 4936 AAGTGCACAAAGAGTCTCTATATACATTAATTAAGTGAAGATTTGACTGAAGTGG 4995
Db 1567 -----ProAsnLysThrTyrThrAsn-----SerAsnLeuGlnLysArgCysGlu----- 1581
QY 4996 CTTTTCCTCTCGTGCAGCTACGTATATGTTTCATGATGATGATGCAATTATACAGAT 5055
Db 1582 CysGlnProProGln-----Glu 1587
QY 5056 CCAAGAGTTAAAGATGAAAATGGCTTCCGAAAAGATTGATGGAAGTGGCGCAACGAA 5115
Db 1588 ProProProGlyProGlnLysGlyAlaArg-----SerAspSer 1600
QY 5116 GGTACAAATTTGGTCACTACAAAGAAAAGAAAAGAAAAGAAAATAAACGCTG 5175
Db 1601 Gly-----ProArgAspThrThrProArgProAlaGlySer 1612
QY 5176 GATGCGCACAATATCTTATGAGGTCCGCTGTAGTGTATGAATATATAGTTTAT 5235
Db 1613 AspAlaArgSerAsnThrValProSerProProArgProAlaGlyAspThrValHis 1632
QY 5236 GATTTAAAGATATATCTAGTATGATTAATTTGAAAGTGAAGAAAAGCAAGCCGAA 5295
Db 1633 GluValAlaGluVal-----GlnGluGluGluGluGluAspGlu 1645
QY 5296 GAAATTTTGAAGAAATATTTAACAAAAATGACATCAGTTGGCAAGGAAGTATAGT 5355
Db 1645 ----- 1645
QY 5356 ACTACAGAAATCCGGTAGTACTGCGGAAAATTTTCTGAGCAAAAATAGAGATGT 5415
Db 1645 ----- 1645
QY 5416 GTGTGAACGAATGATATGGGGTACAAACGCTGTAGGATGATGAATAGTGAAT 5475
Db 1646 -----AspGlyAspLeuProGlu 1651
QY 5476 AGTGCAGAGATGATGATCTTAAAAAATGCTTCTGCTGCTGCTGCTGATGATAT 5535
Db 1652 Asp-----GlnAspGluAspValGluValAlaGluGluGluGluGlu----- 1667
QY 5536 CCTATGGGAAAATGCGGATGAGTACTGCTATACGTTCTTGTGAGGTTTGCCGAA 5595
Db 1668 -----LeuAspValGlyValAla-----ArgIleLeuArgGlyArgThrAsn 1681
QY 5596 TGG-----GTCGACATTTTTCACAAACATAAAGAAATGAGAAATTTGAGGGCG 5652
Db 1682 SerProAspGluAspGluAspGluAlaSerGluGluGlu----- 1694
QY 5653 TGTATGATTAATCTGTGTATATGAAGTAAAGAAAGAAATGACAGATCGGT 5712
Db 1695 -----AspAspAspAspAspAlaGlnAspThrThrGluValThr 1708
QY 5713 ACACATATATATATTTATAGTGTGAACACACAGTATGAAAACAAATCAAAA 5772
Db 1709 GlyGln----- 1710
QY 5773 TATGTGACAAATTAAGACAAAATATATTCGAGCATCTGTGCAAAAGATCAGAGAC 5832
Db 1711 -----GlyGluGluGluThrAlaGluAspHisGlnAsp 1721
QY 5833 GCTCGGATATTTAGACAAACATTTAAAAATTTGTGAAAATTAAGTGAATGT 5892
Db 1722 ThrThrGluGluThrValAspGln-----GluLysAlaGluGluAspGlyGlyGly 1740
QY 5893 GAATATTAAGTATGAAAGATGTCCACACAGCATTAAGTATGATATGTCAAAT 5952
Db 1741 Glu-----ThrProGlnLysGluThr----- 1747
QY 5953 ATGCCCGATCTTAGACAGTGAACCAAAAGTTGAAGAAAGTGAATGTCAAGTG 6012
Db 1748 -----GlnProLysValGluValAsnProCysAsp----- 1757
QY 6013 CCACGAGTCCACGCTGTACGAAGGAAACCCGTCACACGGGTACTGATATCA 6072
Db 1758 -----IleValLys 1760
QY 6073 AAGAGCAGGATGCAAAAAGAAAGCAACAGCCGCTACAAAACAGCCGAAAAA 6132
Db 1761 ThrLeuPheThrThrGluThrLeuLysGluLacysProThrLysTyrValAsnGly 1780
QY 6133 GTGAAAATTTAACAACGAATGCGACACAACACAGACGACGACACAACAA 6192
Db 1781 ArgGlu----- 1782
QY 6193 ACACGAAAACGACATCAACGACAAACAACAGATCTGACGTGGCACAAATGTAAAG 6252
Db 1782 ----- 1782
QY 6253 GCCATTTCTTGAAATTAACCGATACGAGGGGTGAATAGAGGTTTAATCCAAAAACG 6312
Db 1782 ----- 1782
QY 6313 TATGACAAATATCTTAATGCGGTTGTATGCTAGTAACTAAAGAAAATGAATGCG 6372
Db 1783 -----LysPheProAsnThrLysCysIle-----SerSerLysSerAspAlaSerGlySer 1799
QY 6373 ATATGATGCTCTCTAGCAGAAAATTAATGATTAATATATACAAATTAATATAT 6432
Db 1800 IleCysIleProProArgArgGlyLeuThrLeuHisLysIleGluGlyValAspThr 1819
QY 6433 -----GAACTGAAAATTAAGCTGACATATAT-----ATAAAGAGGCT 6471
Db 1820 ThrValSerSerAspGlyGluThrThrProIleThrHisAspAlaLeuArgGluAla 1839
QY 6472 TTTATTAATGTCAGCAATAGAACTCAATTTTGTGCTTAATAATAT-----ATAATGAA 6528
Db 1840 PheIleGlnThrAlaIleValGluThrPhePheLeuThrHisArgTyrLysLysIleLys 1859

Db 2348 LysLysSerProCysIleasnGlyCysGlnAsnLysCysasnCysValGluIleTyrPile 2367
QY 8434 TCTCTATTCGATATACATACAAATTAATACAAAGAAATGTATGACACCAATATATACA 8493
Db 2368 GlnLysLysSerGluIlePglYlyValArgGlnIleTyr----- 2381
QY 8494 AAAATCTCTACTATATCATCATCTGTTCAAAATTTGTACAAAAGTTGAAAACCTTTTAAAGT 8553
Db 2382 ---LleasnGlnIleTyr----- 2385
QY 8554 GAATGTTCTGTGAGAGCTTTTCTGAATATCTTCATGAACAAGTAAGTGTGATTAAT 8613
Db 2385 ----- 2385
QY 8614 AAATTAATGAATAATGATGCTTCTCCATATACGAACATATGCTTCGAAGAACCA 8673
Db 2386 ---ArgAspLysAsnSerAsn-----GlnLlePheGlu-----Val 2396
QY 8674 AAAAGTTATGAAGAGCTTCGATTCATACATACCTTTCAAGATCCATTCGATTAATGCT 8733
Db 2397 LysSerPheLeuGlnIleThrLeuIleProGlnIlePro-----Val 2409
QY 8734 CCTACCGATCAAAACAA---GATGGATGAAGGAATTAACAACTTTTACCTTCGCTG 8790
Db 2410 ValThrAspLysGlyLysHisAspSerLeuThrGlnLeuLysLysLeuLysCysAsn 2429
QY 8791 AAGATGATTAATGATTAATATCTTGAATTAATGGAACGATACCTGTTCTTAATAGTTCA 8850
Db 2430 CysSerGlnLysSerGlnLysn----- 2436
QY 8851 GATGATTAACAAAGCTGATGATTCCTCCACAAAGAACATTTATGTACAAGCCTATC 8910
Db 2437 SerAsnGlnLysAspVal-----Leucys----- 2445
QY 8911 ACTGCATATTAATTAAGAAAGGTGATTAAGAAATTTTAAAAAAAACTTCTACTTCT 8970
Db 2446 ---LeasnLysLys----- 2449
QY 8971 GCTTCACGACGAACAATTTAGTCAAAAATTAATTAATCGAAGAGAGTTGCTTT 9030
Db 2449 ----- 2449
QY 9031 GAGCAATGAATATAGTTATGACATATTCGATATTAATGAAGAACTGATATGATG 9090
Db 2449 ----- 2449
QY 9091 GACACTTATTAATCTGAAAAAATTAATAATTTGAAAACATCAATGACCAACCGAA 9150
Db 2450 ---LeuGlnAspLysAlaLysAsn----- 2456
QY 9151 AATGCTAAACATGCTGGGAAAAATTAATAGCTCAGATATGCGACGCTATGTTATGCGA 9210
Db 2456 ----- 2456
QY 9211 TATAAATTTGCTACTTCAAAAGTATACATTAAGTGAAGATGCTGTCATTAACAAAGAT 9270
Db 2457 ----- 2459
QY 9271 GAAGAACTAATCACTTCTGCTGTTAATGTAATGGCAAGCAAGCATGTAAAGAA 9330
Db 2460 GlnAlaSerGlyLle----- 2464
QY 9331 AAGAAACATGTAAAGTATTCATTAATAAACAATGCTCTGTTCAACGAAGATATTTT 9390
Db 2465 ---ProCysProGlnIleThrSerGlnLysnPro 2474
QY 9391 GAAGGCTCAGATTAATTAAGCAACCTGATGTCAGATGATATAGAAAATATATATGTC 9450
Db 2475 AsnAspGlnAspLleLeuLeuGlnLle----- 2483
QY 9451 TTGAATATATTAATTAATAATACAAATCTAATATATATATATATACCAATTAATA 9510
Db 2484 ---GlnAsnProValGlnLleProAsnIle----- 2492

QY 9511 GATCAATCTTCAGTATATATAGCAATAAACCATCGAAGAAATGTCAGTCATATATA 9570
Db 2493 -----CysProLysValGlnGluIleProGlnProVal 2503
QY 9571 AAATCAAAAGATTCATACATGCGCTTGGAGTTAAATGATATTAAGAAATAGTACAGA 9630
Db 2504 ---GlnGluIleLysCys-----AspLleAlaGlnLleProSerLys 2516
QY 9631 AAAAAAATATGAATAATATGAATTCAAAGAAGTACTAAAAAATATATATCTGTTTA 9690
Db 2517 GluSerSerThrGlnLysnSerGlyLle----- 2526
QY 9691 TATTTGTTGAAGATGAACACAAACAAATCATGTACTAGATGAATATTAAGAGAA 9750
Db 2527 -----GlySerAsnSerGlnGlnAsnProArgSerLysProGln 2539
QY 9751 GAGCAACAGCTTCGCTCAACACTTAATTTCTTAACACCCATGATGATTCCTTCTAT 9810
Db 2540 GlnIleProProProProThrSerSerGlnThrAspThrPro----- 2553
QY 9811 CAGACACCTTTATCTCAACATGACATGACCAATATGAT---CTTAAAT----- 9861
Db 2554 ---ProProAlaProProThrIleGlnProSerGlnAlaAspGlnProThrAsnSerIle 2572
QY 9862 ---GATATATGAAGATGATCTGCTGTTGTTATGATCGCGCTGTTGTTGATACG 9918
Db 2573 SerAspLleLeuSerSerThrIleProPheGlyIleAlaIleAlaLeuThrSerIleVal 2592
QY 9919 CTTCATTCATGACAGAAAAAATTAATCAATGCTGTCGACTGTTGCTGATACGATATC 9978
Db 2593 PheLeuPheLeuLysLysLysThrLysSerSerValAspLeuLeuArgValLeuAsnIle 2612
QY 9979 CCGCAAGAGAGTATGATGCTGCTGTTGATGAATCCAAATATAGTACATACATATGA 10038
Db 2613 ProLysGlyLysIleGlyMetProThrLeuLysSerSerAsnArgLysIleProTyrAla 2632
QY 10039 AGTGTCCATATAAGGCAACATATATATATATGAAGAGATGATGAGAGATGAGAA 10098
Db 2633 SerAspArgLysLysGlyLysThrLysIleThrLysMetGlnGlyAspSer-----AspSer 2650
QY 10099 GATAAATATATGCTGGACTTATCTTCCCTGATATATCTTATCCGAAAGTACATGAA 10158
Db 2651 GlyHisTyrIleGluAsp-----ThrThrAspValThrSerSerGlnLysIleGlu 2668
QY 10159 GAATGATATTAATGATATATATATGATACAGCTGCTCAATATTAACATATGATGAA 10218
Db 2669 GlnLeuAspLleAsnAspLleIleValProIleSerProLysIleThrLeuIleGlu 2688
QY 10219 GTAGTACTAGAACCAATCAAAAGGATATACCAAGTAT-----GATACCACAAGT--- 10269
Db 2689 ValIleGlnIleProSerGlyAsnAsnThrThrAlaSerGlyLysAsnThrProSerAsp 2708
QY 10270 -----AATGATACACCCAGTACAGATATGATATTAATGATGATGAA 10311
Db 2709 ThrGlnAsnAspLleGlnAsnAspGlyLleProSerSerLysIleThrAspAsnGluTyr 2728
QY 10312 AATGAACGAACATGATTTGATTCATATATTAATTAACAAATACAGAACCAAT----- 10365
Db 2729 AsnThrLeuLysAspGlnPheIleSerGlnIleThrLeuGlnIleProAsnAspVal 2748
QY 10366 ---AATATTAACAAAGTGCAGATATTCCAATGAATACAGAACCAATATATATCT 10422
Db 2749 ProAsnAspLysIleThrSerGlyAsnSerSerThrAsnThrAsnIleThrIleThrSerArg 2768
QY 10423 GATAATCTGAAGAAAAACCTTATATATATATATATATATATATATATATATATGCG 10482
Db 2769 HisAsnValGlnGlnLysProPheIleMetSerIleHisAspArgAsnLysIleThrGly 2788
QY 10483 AAGAAATTAATGATATATATATATATATGATGATGATGATGATGATGATGATGAT 10542
Db 2789 GlnIleLysnIleAsnValAsnMet-----ValAsnThrMetAspLleProIleAsn 2807

OY 10543 GCTAGAAATCTTCTTATAGAGTATAGATTATTAATGATTCAGTACTGTTGCTAA 10601
DB 2808 ARGASpAsnAsnValYrSerGlyIleAspLeuIleAsnAspAlaLeuAsnGly----- 2825
OY 10602 CCTATGATATATATGATGAAGTATG 10628
DB 2826 ASPTyrAspIleTyrAspGluValLeu 2834
RESULT 6
O26030
ID 026030 PRELIMINARY; PRT; 3026 AA.
AC 026030:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Variant surface protein (Fragment).
GN VAR.
OS Plasmodium falciparum.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5833;
[1]
SEQUENCE OF 2044-2922 FROM N.A.
RC STRAIN-IT 4/25/5;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatsole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
RA Peterson D.S., Ravelch J.A., Wellens T.E.;
RT "The large diverse gene family var encodes proteins involved in
RT cytoadherence and antigenic variation of Plasmodium falciparum-
RT infected erythrocytes.";
RL Cell 82:89-100(1995).
RN [12]
RP SEQUENCE OF 2044-2922 FROM N.A.
RC STRAIN-IT 4/25/5;
RA Smith J.D., Chitnis C.E., Craig A.G., Roberts D.J.,
RA Hudson-Taylor D.E., Peterson D.S., Pinches R., Newbold C.I.,
RA Miller L.H.;
RT "Switches in the expression of Plasmodium falciparum var genes
RT correlate with changes in antigenic and cytoadherent phenotypes of
RT infected erythrocytes.";
RL Cell 0:0-0(1995).
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN-IT 4/25/5;
RX MEDLINE=99094502;
RA Smith J.D., Kyes S., Craig A.G., Fagan T., Hudson-Taylor D.,
RA Miller L.H., Baruch D.I., Newbold C.I.;
RT "Analysis of adhesive domains from the AAVAR Plasmodium falciparum
RT erythrocyte membrane protein-1 identifies a CD36 binding domain.";
RL Mol. Biochem. Parasitol. 97:133-148(1998).
RN [14]
RP SEQUENCE FROM N.A.
RC STRAIN-IT 4/25/5;
RA Smith J.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RN [15]
RP SEQUENCE FROM N.A.
RC STRAIN-IT 4/25/5;
RA Ryes S., Smith J.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: I42244; AAD03351.1; -
DR InterPro: IPR004258; PFEEMP.
DR Pfam: PF03011; PFEEMP.2.
FT NON_TER 3026 3026
SQ SEQUENCE 3026 AA; EC0A971A82193788 CRC64;

Alignment Scores:
Pred. No.: 3, 29e-194
Score: 4010.00
Percent Similarity: 42.808
Best Local Similarity: 29.768
Query Match: 20, 618
DB: 5
Length: 3026
Matches: 1095
Conservative: 480
Mismatch: 1151
Indels: 954
Gaps: 127

US-10-087-013-1 (1-10628) x Q26030 (1-3026)
OY 88 GAAAGTCACAAAGATGCGACAAAGATGTTGACGTTATGCCAAATATATA---AGCAT 144
DB 13 GUAAspAspIleAspAlaValSHISValLeuAspSerIleGlyGluValITyTyrGlu 32
OY 145 CCATCAAAATATGCAAAAGACATGCGATTCGTTGAAAGGGATTTGACGAAAGAA 204
DB 33 ValIleAspValAlaAlaAsnTyrSerSerGlnLeuAspGlyThrLeuSerAsnAlaIle 52
OY 205 TTTTCGTCGTCGTCCTTCACGCCAGTAATACATATATATATATATATATATATAT 264
DB 53 PheGluAsnGluProIleGlyGlnGlnIleThrGluAsnAsp-----ProCys 67
OY 265 AATTAGATCATAGACAACTACTAATTTACGGTATGATGATGATGATGATGATGATGAT 324
DB 68 LysLeuLeuTyrGluTyrThrIleThrAsnValThrLysGlyHis---GlyArgGlnHisPro 86
OY 325 TGCCATGGGTAGACAAACCAACGATTTGATGAAAGTGAATGCAATGCGAAT--- 381
DB 87 CysArgLysGlyThrGluLysArgPheSerAspValGlyGlyGluCysAspAsnArg 106
OY 382 AAATAGCTAATATATAAAGAAAGAAAGATGATGATGATGATGATGATGATGATGATGAT 441
DB 107 LysIleLysAsp-----SerLysAsnAsnGlyGlyAlaCysAlaProTyrAlaGluLeu 124
OY 442 CATATGTCGATATAAATCTGGAAGCTCTAAAGATTAATATACCAAAATATTCATGAT 501
DB 125 HisLeuGlyValArgAsnLeuGluAsnIleSerAlaLeuAspLysIleAsnAsnAspThr 144
OY 502 TTTATGGGAAATGCTAGTACAGCAAAATACGAAAGTGAATGATGATGATGATGATGAT 561
DB 145 LeuLeuAlaAspValCysLeuAlaAlaLeuHisGlnIleGlnSerIleThrGlnAspTyr 164
OY 562 CCACATGAA-----GAACTTCACAGCGTTCTACTGCTCT 597
DB 165 ProLysTyrGlnAlaGlnTyrAlaSerPheSerProSerGlnIleCysThrMetLeu 184
OY 598 GCACGAGTTTTCAGATATAGTATGATATGTAAGAGATAGATATGTTAAACCAAT 657
DB 185 AlaArgSerPheAlaAspIleGlyAspIleIleGlyLysAspLeuTyrIleGlyAsn 204
OY 658 GTCCATGACAA-----GTAGAAACGGCTCCGAGAGTTCGCAAAATATCATGAT 711
DB 205 LysLysGluLysLeuAspLeuGluLysAsnLeuLysIlePheGlyLysIleTyrGlu 224
OY 712 GGAATGGAAGAT---GAAATGAAATGATTAATGATGATGATGATGATGATGATGATGAT 768
DB 225 LysLeuThrAspProAlaGlyAlaLysAspHisTyrLysAspLeuProAspAsnAspPhe 244
OY 769 AAATTAAGACAGCAGTGGATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 828
DB 245 GlnLeuArgLysAspTyrPheAsnAlaAsnArgGlnIleValTyrPheAlaIleThrCys 264
OY 829 GATGCATCATATTAATGATATTTATGATGATGATGATGATGATGATGATGATGATGAT 888
DB 265 HisAlaGlyLysLeuSerAspLysTyrPheArgLysThrAlaCysGlyThrGlyThr 284
OY 889 AATCTTAATGCGCCATTAACAAAGAAAG-----GTTCCATAC 927
DB 285 GlnGlyLysArgCysAsnAspAspLysLysProGlySerAsnThrAspProPheThr 304
OY 928 AATTAGATTATGCTCCATATATTTACGTTGTTGACGAAATGCGGAGAAAGATTTCG 987
DB 305 TyrPheAspTyrValProGlnTyrLeuArgTyrPheGlnGluTyrPheSerGluAspPheCys 324
OY 988 CGAAAGAAATATTAATGAAAGAGTCAAGACCTCTGCT-----AAT 1035
DB 325 ThrLysArgLysHisLysLysGluAsnAlaIleGluLysCysArgGlyGlnAspGlySer 344
OY 1036 GACAAAGACGCTTATGTTATGATGATGATGATGATGATGATGATGATGATGATGATG 1095
DB 345 GlyLysGluLys---TyrCysAspPheAsnGlnIlePheAspCysLysGlyThrAlaSerGly 363


```

Db 1023 -----gylglulyslleasnCysgly-----AlathrProTyrAspPtyrIle 1037
QY 3142 CCACAAAAAATTAAGATGATGACCGCAATGGCAGCAATGCTACTGCAAGTGCACAAAAAG 3201
Db 1038 ProGlnArgLeuArgTyrMetThrGluThrAlaGluThrPheCysLysGluGlnSerArg 1057
QY 3202 GAGTATGATAAGTTGAAGAGCAAGTGTAAAGAGTGTAAAGCAATTAATGGTCAAGGC 3261
Db 1058 LeuTyrAspGluLeuMetGlyLysCysGlyIleCysMetAsn-----Gly 1072
QY 3262 TGTCAGAAAGAGAGTGTACAGGTTGTACAGTGCACAGCAAGCTGTAAATGATATAT 3321
Db 1073 IleCysAsnLysValLysAspAspCysAlaLysCysThrGluAlaCysLysGluTyrLys 1092
QY 3322 GATATATAGATATTGAAAGAAACAATGCAATTAATTAATCAATTAATATC-----AAA 3375
Db 1093 ThrLysIleGlnProTyrLysAspGlnTyrPheLysLeuGluLeuGluTyrAlaLeuSer 1112
QY 3376 GAATTACATGACACAGCAACAAATGCTGTAAATAGTAAATAGTGTATGAAAGCTTCCAGTACT 3435
Db 1113 TyrLeuHis----- 1115
QY 3436 GCCAAAAATCATATATACAGG-----AATGTT 3462
Db 1116 AlaLysAsnAspSerArgTyrMetAlaPheGlyGlyThrAspProAspTyrGlnGlnVal 1135
QY 3463 ATTGAATTTTGTGCGAATTATACCAACAA-----AATGTTGCGAAAGTAAATAAAGTGT 3519
Db 1136 ValHisPhePheLysGluLeuGlnGlnAlaIleLysSerSerThrSerLysAspProLys 1155
QY 3520 ACTAGTATGAAAGTGTGCTGATGATGCTACACCCACGATGCAAAATGTTGGAGCATAT 3579
Db 1156 ArgSerThrAspAlaIleThrThrAspProThrThrProTyrSerThrAlaAlaGlyTyr 1175
QY 3580 CTCATGATATACAGAAATTTGATGATGTCAGTGCACCAAAATAGTGTGATGATGAA 3639
Db 1176 IleHisGlnGluIleGlyAsnAlaGlyCysGlnIleGlnLysHisPheCysAspAspAsn 1195
QY 3640 AGTAGTATGAAAGTATGCAAAATATGCTTTAGAGATTAACCAACAGACCATGATGCT 3699
Db 1196 LysAsp-----LysTyrValPheArgGluLysProLysAspHisAspGlu 1210
QY 3700 GCGGTGTGTGTAAATGATGCAAAACCGACAGGCTACAGATTAACGAAAAA 3759
Db 1211 AlaLysAsnCysThrGluAsnValGlu-----LysProLys 1222
QY 3760 GCGGAAGAAAAAGATACGAA-----TGTAACAAAGATGATATATCTAAAGAA 3810
Db 1223 LysGlnGluIleAspThrLysLeuAspValCysAlaIleValLysThrAlaLeuThrThr 1242
QY 3811 AACGATGAAAGAAACAAGTAGAGATGTCATCCAAAAAAGAAATAGTAAATGATATCCC 3870
Db 1243 GlnAspAsnLeuThrLysAla-----CysGlnGlnLys-----TyrGlyHisPro 1257
QY 3871 GAT-----TGCATGTC-----GGAATATATAATTAAGTGAAGACCCCT 3909
Db 1258 GlnArgHisThrPheLysPheLysCysIleSerGluThrThrThrLysSerSerAspSerGly 1277
QY 3910 CGTGTGTGTATGCCCCCTAGAACAAAGTATGCGTACATCTTCTGGCAAAAT----- 3963
Db 1278 SerIleCysValProPheArgGlyArgLysLeuTyrValThrProLeuThrLysThrAla 1297
QY 3963 ----- 3963
Db 1298 GluGluAlaThrGluGlyProThrSerProGlnAlaGlyGlyGluAlaThrLeuProThr 1317
QY 3964 -----GATATGAAATATAAAATATACATACACAAAGTAAATTTAAAAAGACT 4011
Db 1318 ProAlaThrAlaSerSerGlnAlaProLysGlyAspSerLeu-----LeuLeuThrAla 1335
QY 4012 TTCATCAATGTGACAGACGAAACATTTCTTCATGATGATTAATTAAGTAAGAT 4071

```

```

Db 1336 PheIleGlnSerAlaAlaValGluThrPhePheLeuThrPheLysTyrLys---MetAsp 1354
QY 4072 GGTGAGCAAAATGACATCGATTAAGAAATTAAGAAAGCAAAATTCCTCCGCAATTTTG 4131
Db 1355 AsnAsnGlyLysAspAlaGlnAspLysLeuLysLysGlyGluIleProGlnGluPheLys 1374
QY 4132 AGATCCATGTTCTCACATTTGGAGATTATAGCAATTTTATTATTGGACAA----- 4182
Db 1375 ArgGlnMetPheTyrThrLeuAlaAspTyrLysArgAspIleCysIleGlyValLysGluAsp 1394
QY 4183 -----GATATCAAAAGGTGATGCTGAGGAGTAACTAAAGTAAAGAG--- 4224
Db 1395 ValIleLysAlaLeuGluAlaSerSerAspAsnLysSerGlyAsnAlaIleLysGluIle 1414
QY 4225 -----CAATAGATTTCTTTTCAAAAATGTCAGCCAAAATTCCTTAATGGA----- 4272
Db 1415 SerAspLysIleGluAlaIleLeuLysGlnSerGlySerLysLeuProGlyGlyLeuPro 1434
QY 4273 -----AAACACGCCCAAGATGCTGGACAGAACATAGTCATGAGATA 4314
Db 1435 ValThrProAsnAsnValLysAsnProLysThrThrPheAspGlnAsnAlaLysHisIle 1454
QY 4315 TGGGAGCTATGCTATGTCAGTACTGA-----AAATTTGGGCGCAAAA---AAA 4359
Db 1455 ThrHisGlyMetValCysAlaLeuThrTyrLysGluAspThrGlyAlaLysGlyLys 1474
QY 4360 GATGATTTTACCGAAAC----- 4377
Db 1475 ThrSerIleThrGlnAspProThrAlaTyrGlyLysLeuTyrAspAsnAspGlyLysLys 1494
QY 4378 -----TACGTTACACACAGCTCAAAATTAAGT----- 4404
Db 1495 ProLysGluAspLysTyrAspTyrLysIleValLysIleSerSerValProSerSerAsn 1514
QY 4405 -----GACAAACACACACACTTGGAGAAATTTGCAAAACACCCAGCTTTTACGATGG 4458
Db 1515 ValProSerGlyAspThrLysLeuGlnLysPheSerArgAspProThrPhePheArgTyr 1534
QY 4459 CTAAACGAATGATGACGACGACTATGCTATACAGACACAAAAATTTGAAAGATGTGACG 4518
Db 1535 LeuGlnGluTyrGlyGlnGluPheCysArgLysArgLysAspLysLeuGluLysIleGlu 1554
QY 4519 GAAAAATGTAAG-----TCAATGACCAATTG 4555
Db 1555 LysGlnLysTyrGlyLeuAsnTyrArgLysArgLysIleTyrCysSerGlyAspGlyHis 1574
QY 4546 AAGTGTGAT-----ACGAA 4560
Db 1575 IleCysGluLysThrAspThrSerArgAsnAsnThrPheIleAspLeuHisCysProArg 1594
QY 4561 TGTAATAGAAATGCGAGGACTACGTTAAATATATG-----AAAAAAGAGTGTGAT 4617
Db 1595 CysLeuLysGlnCysIleLysTyrLysArgTyrPheGlnLysGlyGluPheHis 1614
QY 4618 CCAAGATTAATTTACAGAGATGAACGCGACAAAGAAATTCGATGACACACAT 4677
Db 1615 AsnGlnLysAsnAsnTyrGlyLysGlnPheAsnAsn-----AspLeuLysGluLys 1631
QY 4678 GGTGTAATGCTTACAGACTATATGGAAGCAAGATCAACAGATCTTGAACAGAAAAAT 4737
Db 1632 Gly-----TyrSerPheAsn-----AspPhe 1659
QY 4738 ACTGCTAGTTGTGTGAT---AAGCCTGGAAGTCCCTGTGTGTAAGAAAT---ATA 4791
Db 1640 LeuAlaSerLeuAsnHisCysLysHisGlyLysHisAsnArgAspLysAsnLysIle 1659
QY 4792 CAATTGTAGAAAAACAGGCTTACTATGATGCCAGCAAACTGT---GGTGCACAAA 4848
Db 1660 GluPheAsnAsnHisThrLysThrPheGlyProSerGluTyrCysLysAlaCysProVal 1679
QY 4849 TTATTTGCAATTCAGACAAATATATCAATTTTCAAGTAAAGATGAAGTAAAGAT 4908
Db 1680 TyrGlyValLysCysAsnLys-----LysAsnGlyGluCysGluThrIle 1694

```

QY	5923	CAGGATTAACCTAGTGAATAGTCAAAATATGCCCGATCATTAAGCATGACCAAA	5981
Db	1971	ThrlYsLthr-----	1976
QY	5983	GAAATTGAAGAAAGTGAATTTTCAGAGTGCACAGAGTCCACACAGCTGACGAAGGAA	6041
Db	1979	-----	1984
QY	6043	ACAACCTCACACAGGGTATACACTGATATCAAAACGACGCGCATCGAAAAAGACCGAAA	6107
Db	1984	-----	1984
QY	6103	ACAGCGCGCGCTACAAAAACGCCGAAAAAGTGGAAAAATCTAACACAGAAATGGAGACA	6162
Db	1984	-----	1984
QY	6163	CAAAACAGAACCCGACGACGACACACAAACAGAAAAAGCAATCAACAGCAACAAACA	6222
Db	1984	-----	1984
QY	6223	ACAGATTCGACGTGGGACACATGGTAAAGCCATCTTTGCAATTAACCAATGACGAG	6282
Db	1985	-----	2000
QY	6283	GGTGGATAGAGGGTGTGTAATCCAAACGAT-----GGAACAAATCTCTAAATGG	6333
Db	2001	LysLysLleGluGluCysAsnThrLysTyrLysProThrLysAsnAspLysProGlyTrp	2020
QY	6334	GGTGTATGTAGTACGTCAAAAGAAATGAAATAGGCATGTGATGCTCTTAGAGAGA	6393
Db	2021	AsnGly-----ThraSpLysValIleAsnAlaGlyGluGlySerCysMetProProArgTrg	2039
QY	6394	AAAAAATTGTATTAATTAATATATATCAATATTAAATTAAATCAAGAAATTAAGGTGAC	6453
Db	2040	GlnLysLeuCysIleHisAsnLeuGluHisLeu-----SerGlnLysAlaThrGlu	2056
QY	6454	AATGATATTAAGAGCGCTTATTAATTAATGTCGACGAATGAACATTTTGTGTTA	6513
Db	2057	ThrGluLeuAlaGlyLysAlaPheIleGluCysAlaAlaIleGluThrPheTrpLeuTrpAsp	2076
QY	6514	AAATAT-----	6531
Db	2077	LysTyrLysGluAspLysLysAspGluLysLysThrGluGlyGlyIleSerAspAsp	2096
QY	6532	CCT---GCAGACAGAAATGCAATTCGCAAAATGCACAAATCCAGATGAATTTAAAGATA	6588
Db	2097	ProAspAspProGlnLysLysLeuGluGlyThrIleProGluAspPheLysArgGln	2116
QY	6589	ATGTATTATACATAGTGGTGAATTAAGAAGATAGTTTGGAAACGATTTCTTAATGAT	6648
Db	2117	MetPheTrpThrTyrGlyAspLysTrpAlaGspPheLeuPheGlyThrAspIleSerLysGly	2136
QY	6649	AAA---AAATATTACTGTATACAAAAATAGTGTACAAACATTCCTCAATGAATAATTAAG	6705
Db	2137	HisGlyLysGluSerAlaLeuGlyLysLysIleAspSerLeuPheLysAsnGlyAspGln	2156
QY	6706	AAAAAACGATTAATAAAAAAGATGAAGATTAAGTAAATATTTTGGGAGAAAAATAAA	6765
Db	2157	LysSerProSerGlyLysThrProThrGln-----TrpTrpAsnAspTyrGly	2172
QY	6766	AAATTTATTTGGAGAGAAATGATATGATTAATCAATCACTCTCACAGAGCAAAAGCA	6825
Db	2173	ProAspIleTrpLysGlyMetValCysGlyLeuSerHisIleLysAsnGlyAsn---	2191
QY	6826	AAAAAATAATTAAGATTAATAC---CAGTACAAATGACATGACCAAACTGACGCTTCC	6882
Db	2192	LysLysGluGlnLeuArgLysAsnLeuThrAspAsnAsnLysTyrThrLysIleSerSerLys	2211
QY	6883	CTTGAAAGATTTGTAATAAAGCCCAATTTTGGATGCGTTCCACAGAAATGGCAGAAAGA	6942
Db	2212	LeuGluAspPheAlaSerArgProGlnPheLeuAlaTyrIleGluTyrIleGlyAspGln	2231
QY	6943	TTTTTGTATTAAGAGCAAGCAAGTGTGTTAAATTTGAGAGCGGCTGTATAAGAAATTGAG	7002

```

Db 2232 pheCysAtGclUatGValValLysIleasnInLeuLysThrGlyCysasnGlyTgU 2251
OY 7003 TGTATGTAGTATGATGACGCTGATGACAGAAATGTGCAGAGCGCTGTGTACATCTTCAA 7062
Db 2252 CysGlySerGlnLusnGlyLysGluAlaCylLysAsnLacysGluAlaTyrIys 2271
OY 7063 AATTTATTTAGAGAGTGAAGAAATGTAATGAAAGACAAAGAAAGTTCAAAAGAT 7122
Db 2272 SerTrpLeuLysAspTrpCysArgInTyrGluGlnGlnThrAlaLysPheAsp 2291
OY 7123 AAGATGGCAAAAGTATGAGTATTCCTTCACGAAAGACATGAGAAAGCAACA 7182
Db 2292 LysLysAspLysLysPhe--AspGlyThrSerAlaGluAlaValAlaValSer 2310
OY 7183 TGTGTCATGATATTTTAAACATGAAATTTAAAGATTTATGTGCATATAGATTTGCT 7242
Db 2311 SerValHisGluTyrLeuGlnGlnGluLysAsnLeuLysThrLysGluAspCysAla 2330
OY 7243 TGTATGCAAAACCTTCTTCACACACTACCAAAACACACACATCACATCCGAT 7302
Db 2331 CysMetGluLysProSerIleGln-----AspGluGlnLysThrGluLeuGly 2346
OY 7303 GCTAATGATATGCGCAATCGCTGATTTATGTTCTGAGAAATTT--AACAAGTGGAG 7359
Db 2347 GlyAsnTyrPheProGluAlaMetAspTyrProTolysGluLysGlyCysLys 2366
OY 7360 TGTCTGAACTTTCAAAAAAGGATCTATGATTCATACAAAAAATTTACTGACATAA 7419
Db 2367 Cys-----AlaIleProSer 2371
OY 7420 ATTCCTATGATTTGTAGAGAAAGACATATTTATTTCTAAGACAGAAATTAAT 7479
Db 2372 GluProMetSerCysValGluGlnIleAlaLysHisLeuArgGluLysAlaGluLysAsn 2391
OY 7480 ATGATATTAACCTTGAAGAAATTTATACCTATGAGCTGACAAAGAAAGAAAGT 7539
Db 2392 ValLysIleGlyGluSer-----LeuLysGlyThrProAlaLys--Ser 2406
OY 7540 AAAATAGTTGACATATATATATATCTTGCGATCTAAGAAACCTTATGCACTGATAA 7599
Db 2407 LysAsnAspCysThrLysIleAsp-----GluAlaIleLysGluLysAsnGlySerLys 2424
OY 7600 TATATGGAAGA-----AGAAACCTTGTGAAATATAGAGAAAG 7638
Db 2425 IleIleAsnLysSerIleLeuAspSerThrPheAlaSerAsnGlyGlnGlnLys 2444
OY 7639 -----AATGCTTTTAAAGTATGATGAAATGTTACAAAATTCGAAGTTC 7689
Db 2445 AspAlaThrAspArgLeuLysIleGlyLysGlnTrpGlnPhe-----AsnLysIle 2461
OY 7690 TATCAGAGAAAAAGATATGTACTCTCAAGAAAGAACATATGCTTAAGGAAT 7749
Db 2462 AsnGlyThrGlnThrLysLeuTyrValProProArgLysAspPheCysPheAsnAsp 2481
OY 7750 TTTAGATGAATTAATTTGAAAGACTTAAGAGATTAATATCTCTTAATAATGTTTCT 7809
Db 2482 LeuLysAsnIleGlnPheAsnGluValGlnAspSerAsnSerLeuGlnLysIleGln 2501
OY 7810 CGAATCGCAAGAAATGAGAAATAGACATATATAAAACCTCAACACAGAGCGGTCG 7869
Db 2502 HisValAlaLysAsnGluGlyIleAspIleLeuLysLysLeuAsnProGlnAspGlnAsn 2521
OY 7870 GCAATGATTCATATGATGATCTATGAAATATAGTTTGGCTGATCGGTGACATAGTT 7929
Db 2522 AlaPheSerGluLysCysAspAlaMetLysTyrSerPheAlaAspLeuGlyAspIleIle 2541
OY 7930 AGAGGAACAGATATGTTAGCAATGTGCTGTTACTTACCTCCCTACAAATATAATATAT 7989
Db 2542 ArgGlyArgSerLysIleAsp-----ProThrAsnAsnAsnLysIleGlu 2556
OY 7990 AAGCTTTTGAATATATATATGAAATGAGAAATGAGAAATATAAGAGTGAATTAATATAC 8049

```

```

Db 2557 LysGluLeuGlnLysIlePheLysGlnIleGlnAspAspAsnLaserLeuSerLysMet 2576
OY 8050 AAGCATGTACAAACGTTTCTGCTTGTGCTGATGATGCTAATGAAAGATATTTGGAAA 8109
Db 2577 ---GluLeuProGluLeuArgGluLysTrpTrpAspAlaAsnArgLysGluValTrpAsn 2595
OY 8110 GCAATGACGTGCAAAAGCCACGAGAAATGCAAAACCTTTTGAAGAAGAGATGATGCA 8169
Db 2596 AlaMetThrCysValAlaProAsnAspAlaHisLeuLysLysLysAsnAsnProGly 2615
OY 8170 TTTGAGCGCATATCATATACAAAGAT-----AAGTGTGACATTAAGGAC 8214
Db 2616 AsnLysSerGlnIleIleAlaSerGlnThrGluGlnThrLysLysCysSerHisAspSer 2635
OY 8215 GATCCACCTGTGTGATGATTTATATACCTCAACGGTGTGATGATGATGATGATGATGAA 8274
Db 2636 GluProProAspTyrAspTyrIleProGlnArgTyrArgPheLeuGlnGlnTrpSerGlu 2655
OY 8275 TATTTATTTGTAACCACTGATGAGAAATTTGAAAAATTTTAAAAATCATGTGATCAGT 8334
Db 2656 TyrTyrCysLysAlaLeuLysGluLysAsnAspGluMetLysAsnAspCysSerLysCys 2675
OY 8335 ---AAACATCTGACAGATGCAAGAAATGATTTATGAGAAATTAAGTGTGAACAGTGA 8391
Db 2676 IleLysSerGlyAlaThrCysGluLysGluGlnAspLysGluLysCysLysGluCysAsn 2695
OY 8392 ACGAGATGTCAGAAATATATAAAATTTGTTCTTAATGGAATCTGATTCGATATACAA 8451
Db 2696 AspLysCysLysGluTyrLysAsnIleValAlaAspLysTrpLysLeuGlnGln 2715
OY 8452 TCAATTAATACAAAGAAATGTATGAAACAAACATATATACAAAATCTTACTTATGAT 8511
Db 2716 AsnGlnLeuLysLysLysLeuTyrThrGlnAspArgThrHisGlyProSerThrAlaArg 2735
OY 8512 CATGTTCAAAAT-----TTGTACAAAGTGTAAAACTTTAAAGTGAATGTTCT 8562
Db 2736 ArgAsnProSerIleGluLysPheThrGlnLysLeu-----GluLysSerCysAsnAsp 2752
OY 8563 GTTAGAGCTTTTCTGAAATCTTCATGAAACAAAGTAAAGTGTGATTAATTAATTAAT 8622
Db 2753 ProTyrSerLysAspLysLysLysLeuAspIleSerThrIleCysThrAspLysPheSer 2772
OY 8623 GAAATGATGCTTTCATATATACGAACATATGCTTTGCAAGAAACCCAAAAAGTTAT 8682
Db 2773 GluThrAsp-----SerArgGluSerAsnTyrAlaPheSerProTyrProLysAspTyr 2790
OY 8683 AAAGAAGCTGCAAGTGTACACTACCTTCAAGATTCATGATGATTAATGCTTACCGAT 8742
Db 2791 LysGluAsnCysLysCysLysVal-----AsnThrProThr-- 2802
OY 8743 CAAGAACAAAGATGATGAAGAAATTAACAACCTTACCTTCTGCTGAGAAATGATTAAT 8802
Db 2803 -----SerAsnAsnAsp-- 2806
OY 8803 GATTAATATCTTATATTAATGAAACGCAATACCTTGTCTTAATAGTTACAGATGAACAA 8862
Db 2806 ----- 2806
OY 8863 GGTGATATGATTCCTCCAGAAAGACATTTATGTACAGACCTATACCTGATATTAAT 8922
Db 2807 -----ProLys 2808
OY 8923 TATAGAAAAGTATGAAGAAATTTTAAAAAAACCTTACTTCTGCTTCACTGCA 8982
Db 2809 -----Ser 2809
OY 8983 GACATATGTTAGTCAAAAATATAATGGAAGAGAGTGTGCTTTAGGCAATGAAA 9042
Db 2810 ProSerLeuLeuGlyProSerPhe----- 2817
OY 9043 TATAGTTATGACATATATTCGATATATTAATTAAGAAACGATATGATGACACTTCATTA 9102
Db 2817 ----- 2817

```

```

QY 9103 TCTGAAAAATTTAAAAATATTTGAAACATCAATGAGCAACGAAATCTGTAACA 9162
Db 2817 -----
QY 9163 TGGTGGAAAAATATAGACGTCAGATATGCGACGCTATGTTATGTGATATAAATTCGT 2817
Db 2817 -----
QY 9223 ACTTCAAAAGTAACTAGATGAGAGATGTGCAATTACCAAGAGATGAAGAACTAAT 9282
Db 2818 -----
QY 9283 CAGTTTCCTGCTGTTAATGATGCGCAACGACAGCTAAGAAAGAAACATGTA 9342
Db 2821 -----
QY 9343 AGTGATTCATTAAACAAATGCTCGTTCAAAAGACAGTAAATTTGAAGCGTCGAA 9402
Db 2822 -----
QY 9403 TTATTAGACACCTGATGTCAGATGATATGTAAGAAATATATGCTTGAATATATG 9462
Db 2825 -----
QY 9463 ATAAAAATATCAATGAAATCTAAATATATAATATAGCAATTAAGATCAATCTCA 9522
Db 2835 ValLeuHisProPheIleAsnMet-----
QY 9523 GGTAAATATAGACAAATAAACCATGTGAGAAATGTTCAATATATTAATCAAAAGAT 9582
Db 2842 -----
QY 9583 TCTCAATGCGCTTGGAGTTAAATGATATTAATGAAATGTTACAGCAACAAATATAT 9642
Db 2843 -----
QY 9643 GAAAAATATGAAATTCAAAGAGACTAAATAATATATATATATATTTGTTGA 9702
Db 2860 AlaValProGlnPhe-----
QY 9703 GATGAACACACAAATATCATGTACTAGTGAATATTAAGAAAGACGCAACAGTT 9762
Db 2865 -----
QY 9763 CGTCTAAAGCACTATTTCTTTACACCCAGTGAATCTTTCTATCAACACCTTCA 9822
Db 2874 AlaProPro-----
QY 9823 TTCTCAACACATGAGTAGACACAAATATGATCTTAAATATATGTAAGAGTATATC 9882
Db 2881 -----
QY 9883 TCTGTTGTTATGTAATCGCGCTAGTGTGATGAGCGCTTCAATTCATGTAAGAAATATC 9942
Db 2890 ProValGlyIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 2909
QY 9943 AAATGCTGTGAGACTTGTGCTGATGATGATATCCCGCAAGAGATGATGAGATGCT 10002
Db 2910 LysHisProValAspLeuPheSerValIleAsnIleProLysSerAspTyrAspIlePro 2929
QY 10003 ACGTGAATCCAAATAGATGATACATACATATAGAGTGTCCATATTAAGCAAAA 10059
Db 2930 ThrLysLeuSerProAsnArgTyrIleProTyrIleSerGlyLysTyrArgGlyAsnGly 2949
QY 10060 ACATATATATATATGAA-----CGAGATATCTAGTGTGAGATGAAAGTAAATATATGCG 10113
Db 2950 ThrPheThrLeuLysGluIleValGlyThrAspSerGlyTyrThrAspHisTyr----- 2967
QY 10114 GACTTATCTCTCTGATATATCT-----TCATCGAAAGAGATGATGAAGATGATGATAT 10170
Db 2968 -----SerAspIleThrSerSerSerGluSerGluTyrGluGluLeuAspIle 2983

```

```

QY 10171 AATGATATATATCTACCAGTACTCCTAAATATATAAATGATAGACTAGTACTAGAA 10230
Db 2984 AsnAspIleTyrHisHisValLeuLeuAsnIleLysThrLeuIleGluValLeuGlu 3003
QY 10231 CCATCAAAAAGG-----GATATACCAAGGAT-----GATACCAAGTAAAT 10272
Db 3004 ProSerGlyLysLeuSerGlyLysAsnThrIleProThrSerGlyLysAsnThrProSerAsp 3023

RESULT 7
ID Q26031 PRELIMINARY; PRT: 3078 AA.
AC Q26031;
DT 01-NOV-1996 (TREMBLREL. 01, Created)
DT 01-NOV-1996 (TREMBLREL. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLREL. 19, Last annotation update)
DE Variant-specific surface protein.
GN VAR-1.
OS Plasmodium falciparum (isolate Dd2).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=57267;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DD2;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
RA Peterson D.S., Ravetch J.A., Wellens T.E.;
RT "The large diverse gene family var encodes proteins involved in
RT cytoadherence and antigenic variation of Plasmodium falciparum-
RL infected erythrocytes."
RL Cell 82:89-100(1995).
DR EMBL; L40608; AAA75396.1;
DR InterPro; IPR004258; PEMP.
DR Pfam; PF03011; PEMP; 2.
SQ SEQUENCE 3078 AA: 349297 MW: C8037C2BC3CD7G3 CRC64;

Alignment Scores:
Pred. No.: 1,36e-168 Length: 3078
Score: 3501.00 Matches: 1057
Percent Similarity: 38.65% Conservative: 470
Best Local Similarity: 26.75% Mismatches: 934
Query Match: 18.00% Indels: 1490
Db: 5 Gaps: 147

US-10-087-013-1 (1-10628) x Q26031 (1-3078)
QY 25 ATATATAAATGAGGATGCGAGCATCATCTAGAGGAGATCTAAACCCCTATATA 84
Db 4 LeuAlaLysMetGlyProLysGluAlaIleGlyLysAspIleGluLysP----- 20
QY 85 AAAGAAAGTCAAAAAGTGCAGAAATGTTTGGAACTGATGCCAAATATATAGACAT 144
Db 21 -----GluSerAlaLysHisMetPheAspArgIleGlyLysAspValTyrAsp 36
QY 145 CCATCAAAA-----TATGCAAAAAGCAATGTCGATTCCTTAAAGGGATTTGACGAAGCA 201
Db 37 LysValLysGluGluIleAlaLysGluArgGlyLysGlyLeuGlnGlyArgLeuSerGluAla 56
QY 202 GAATTCGCTGCTGCTCTCTACGCCAGTAAATTAAGCATATATATATATATCAATATCA 261
Db 57 LysPheGlyLysAsnGluSerAspProGlnThrProGluLysP-----Pro 71
QY 262 TGTAAATTTAGATCATATAGACATATCTAATTTACGATGATGATGATGATGATGATGAT 321
Db 72 CysAspLeuAspHisLysTyrHisThrAsnVal-----ThrThrAsnValIleAsn 88
QY 322 CCTTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 378
Db 89 ProCysAlaAspArgSerAspValArgPheSerAspGluTyrGlyGlyGlnLysThrHis 108
QY 379 AATTAATATCGTAATATTAATAAAGAAATGATGATGATGATGATGATGATGATGATGAT 438
Db 109 AsnArgIleLysAspSerGlnGlnGlnLysAspAsnLysGlyAlaCysAlaIleProTyrArgArg 128

```

OY	439	CGACATATGCTGTATAAAACCTTGAAGAGCTGTAAATGATATATAATCAAAATATTCAT	498
Db	129	LeuHisValCysAspGlnAsnLeuGluGlnIleGluProIleLeuIleThrAsnThrHis	148
OY	499	GATTATTTGGGAATGACTAGTATACAGCAAAATACGAAGTGATCAATTTAAATAT	558
Db	149	AsnLeuValAspValCysMetAlaIleValPheGluGlnSerIleThrGlnAsp	168
OY	559	CATCAAC--CATAAAGCACT-----TCAGACGCTTGACTGCTCTTCCA	600
Db	169	TyrProLysTyrGlnIleAlaThrTyrGlyAspSerProSerGlnIleCysThrMetLeuAla	188
OY	601	CGAAGTTTGCAGATATAGGTGATATATGTAAGAGCAATAGATATGTTTAAACCAATGTC	660
Db	189	ArgSerPheAlaAspIleGlyAspIleValAlaArgLysPheLysIleGlyAsnPro	208
OY	661	CATCAAC-----AAAGTAGAAGAGGGCTCCGACAGGTTTCAAGAAATA	705
Db	209	GlnGluIleLeuSerGlnArgGlnGlnLeuGlnAsnAsnLeuLysThrIlePheGlyLysIle	228
OY	706	CATGAT-----GGATGGAAGATGGAAGTAAAAAATGATTACATCTGATGATCT	756
Db	229	TyrGluLysLeuAsnGlyAlaGlnAlaArgTyrGlyAsnAsp-----ProGlu-----	244
OY	757	GGAATATATATATTAATTAGACAGACATGCTGGATGTCATATAGAAATTAAGTATGGCA	816
Db	245	-----PhePheLysLeuAlaArgGlnAspTrpTrpThrAlaAsnArgGluThrValTrpLys	262
OY	817	GCTATTAACATGTGATGCATCATATTAATCGATTTTATTCATATCAAGAAAGTAATCA	876
Db	263	AlaIleThrCysAsnAla--TrrGlyAsnThrTyrPhe--HisAlaThrCysAsnArg	280
OY	877	CCATATATTTCAATCCTTAATCCTTAATCCGCGCCATTAACAAGAAAGGTTCCTTCAATTTAGAT	936
Db	281	GlyGluArgThrLysGlyTyrCysArgCysAsnAspAspGlnValProThrTyrPheAsp	300
OY	937	TATGTCCTCATATTTAGCTTGTTGTCGACGAATGGGGAAGAGCTTTGCCGAAAAAGA	996
Db	301	TyrValProGlnIleLysLeuAlaGlyTrpPheGluGlnIleTrpAlaGlnAspPheCysArgLysLys	320
OY	997	AATATTAATTAAGAAAGAGTCAAGACGACTCCTGCTCGT-----AATGACAAACA	1044
Db	321	AsnLysLysIleLysAspValLysAlaGlnCysArgGlyLysAspLysGluLysAsp	340
OY	1045	CGCTTATATTTGTAGCTCATATGAGCATGATTTGTACACACATTTGGAAAAAGCATATT	1104
Db	341	Arg---TyrCysSerArgAsnGlyTyrAspCysGluLysThrLysArgAlaIleLysLys	359
OY	1105	TTCGATTTGGATTAATTAAGTCTACTGCTGCTGACTAAATGCAAAAGTTTGAAGCTTGG	1164
Db	360	LeuArgTyrGlyLysGlnCysIleSerCysLeuTyrAlaCysAsnProTyrValAspTrp	379
OY	1165	TTTAGGATCAACAAGACGATTTTAAAAAAACAAGAAATATGAAAAAGAAATTAACA	1224
Db	380	IleAsnAsnGlnLysGlnGlnPheAspLysGlnLysLysTyrAspGluIleLys	399
OY	1225	TCATAT-----TTATCAACAGATAACAATTTGTCAT	1257
Db	400	LysTyrGlnAsnGlyAlaSerGlyLysSerArgGlnLysArgAspAlaGlyGlyLysThr	419
OY	1258	AATATTAATAGTGAA--TATTAATAACAATTTATGAAAAAATTAGGAAACGCATAT	1314
Db	420	ThrThrAsnTyrAspGlyTyrGluLysLysPheTyrAspGluLeuAsnSerGlnLys	439
OY	1315	GCACACTATGACCTTTTAAATTTACTTAATGAAGAAGAAAGTAAATTCG-----	1362
Db	440	ArgThrValAspLysPheLeuGlnLysLeuSerAsnGlnGluIleLysThrLysValLys	459
OY	1363	-----AAAGAGGATTAACACAGAGAAAGAGATATTACTTTACTACAGCTCT-----	1410
Db	460	AspGluGlnGlyGlyThrIleAspPheLysAsnValAsnSerAspSerThrSerGlyAla	479
OY	1411	-----GATGACAAAGGAGATTTTATCGTTCAAAATATTGCCAAGCTGCT	1455

Db	480	SerGlyThrAsnValGluSerGlnGlyThrPheTyArgSerLysTyCysGlnProCys	499
Oy	1456	CCCAGCTCGCGGGGTCAAAATGTCATGCTATTAATAATACACACAC---AAATCGATTAATCAT	1512
Db	500	ProTyrcysGlnValLys-----LysValAsnAsnGlyLysSerSerAsnLys	515
Oy	1513	CGTGAACGTGTAATAAT-----GAAGACTATAAACCTCCATGGGGT	1554
Db	516	TrpLysGlnLysAsnAsnGlyLysCysLysSerGlyLysLeuTyGlnProLysProAsp	535
Oy	1555	GTGAAGCCTACTAATATACACTGTCCTTTATAGCGTAATGAACAAAGGATATTATACAA	1614
Db	536	LysGlnGlyThrThrLleIleThrLleLeuLysSerGlyLysGlyHisAspAspLleLysLys	555
Oy	1615	AAATTAAGAAAATTTTCTTAAC-----AGCTCA	1641
Db	556	LysLeuAsnLysPheCysAspGlnLysAsnGlyAspThrLleAsnSerGlyGlySerGly	575
Oy	1642	ACTAATTTACAAAGATATAATAAT-----CAAAAATGGGAATGC	1680
Db	576	ThrGlySerGlySerGlyGlyLysAsnSerGlyArgGlnGlyLeuTyGlnGlyLysCys	595
Oy	1681	TATTTATAGGATGCAAAATATTAATTAAGATGTAACCTGGACAAAAATCTGAAATCAATAT	1740
Db	596	---TyrcysGlnLysAspValValLysValGlyHisAspGlnAspAspGlnLysAspTyrc	614
Oy	1741	GATATATCTTAAG-----ATAATA	1758
Db	615	GlnAsnValLysAsnAlaGlyGlyLeuCysLleLeuLysAsnGlnLysLysAsnLysGln	634
Oy	1759	-----TCATTTCTATTAATTTT	1773
Db	635	GlnGlyGlyAsnThrSerGlyLysGlnProAspGlnLleGlnLysThrPheAsnProhe	654
Oy	1774	TTTCAATTAATGCGGTACATATTTATTAAGGATACACTAATTAATGCAATGCAAACTTAA	1833
Db	655	PheTyrcTyrcTrpValAlaHisMetLeuLysAspSerLleHisTrpLysLysLysLeuGln	674
Oy	1834	ACTTGTA---AATTAATTAACAACAGCAGATTT---ATTGAGTAATGTAACGAAATTCG	1887
Db	675	ArgCysLeuGlnAsnGlyAsnAlaGlyLleLysCysGlyHisAsnLysCysAsnAsnAspCys	694
Oy	1888	TTATGTTTTCACAGATGGGTAAACAAAGAAAGAAATGGAATAGTATAAGAACTG	1947
Db	695	GlnCysPheLysAspTrpLleThrGlnLysLysAspGlnUtrpGlySerLleValGlnHis	714
Oy	1948	TTTCACAAAAAAGATATATACAG-----	1971
Db	715	Phe---LysThrGlnAsnLleLysGlyArgGlySerAspAsnThrAlaGlnLeuIle	733
Oy	1972	-----CAATGTAATTATAGTAATATTAATATCTTTTGAGCGTTATTTT-----	2016
Db	734	ProPheAspHisAspTyrcValLeuGlnTyrcAsnLeuGlnGlnLysPheLeuLysGlyAsp	753
Oy	2017	-----TTTAAAGTTATGAGATTAACCTTGACAAAGATGACAAAGAAATGG	2058
Db	754	SerGlnAspAlaSerGlnGlyLysSerGlnLysSerLeuAspAlaGlnLysLysA-----	771
Oy	2059	AAAGACTTATGAAAAATATTAAGAAAAAATAATGACTTTTCCAAATTTGGAAATAAT	2118
Db	772	-----GlnGlyLeuLysHisLysLeuArgLysLleIleGlnSerGlnLysAspAsn	787
Oy	2119	AGGAGC-----TATTAGACATAGCAATACAACTCTTGTA	2154
Db	788	GlnGlnAlaSerValGlyGlyGlyValThrGlnGlnLysAsnLleMetAspLysLeuLeu	807
Oy	2155	GATCACTTAAGAAAGAACTGCCAGCATATGT-----	2184
Db	808	AsnTyrcGlnLysAspGlnLysAspLeuCysLysLleHisGlnAspGlnGlnGlnLysLys	827
Oy	2185	---AAAGCAATATATACAAACGAAGCATGTGGAAACATCCCATTAATGCAACAAACCCG	2241

OY	3295	TGCACAGAGCCTGTATGAATATATGATATATAGATTTAGCAAGAAACATGAAAT	3354
		: : :	
Db	1180	CysLysAlaIlaCysAspLysTyrLysGluGluIleGluLysTrpAsnGluGlnTrpArg	1199
OY	3355	ATATATATCAGATTAATATACAAAGATTACATGACACACACCAAAATGCTGTATAGTAATAGT	3414
Db	1200	LysIleSerAspLysTyrAsnIleLysIleLysGluIleLysThrIleSerThrAsnPro	1219
OY	3415	GCGATTTGAAGCTCCAGACACGCCAAATATCATATGAC-----AGGAATGTTAT	3465
Db	1220	Gly-----ArgThrValLeuGluLysAspAspArgProAspTyrGlnIleMetVal	1233
OY	3466	GAAATTTTGTGGAAATTATACCAACAAATGTGGCAAA-----AGTATAAA	3513
Db	1235	AspPheLeuThrProIleHisLysAlaSerIleAlaAlaArgValIleuValLysArgAla	1255
OY	3514	AGTGGTACTAGATGATGAAGTGGCTGATTTGGTACATACACCAACGATGATAAATGTTGGA	3573
Db	1256	AlaGlySerProThrGluIleAlaIleAlaIleAlaProIleThrPheThrAlaIle	1275
OY	3574	GCAATCTCCATGATATACAGAAATTTTATGATGATTGCATGCCATCAAAATACGTTTGAT	3633
Db	1276	GlyTyrIleHisGlnIleGluTyrGluGlyCysGluGluGlnIleThrGlnPheCysGlu	1295
OY	3634	GAAAA-----AGGATGATGATGATATACGAAAAATATGCTTTAGA	3675
Db	1296	LysLysHisGlyAlaIleThrSerThrIleSerThrLysGluAsnLysGluIleThrPheLys	1315
OY	3676	GATTAACCAACAGACATGATGCTGCTGCTGTTGTAAGAAAGTGATGCAAAACCCAGAG	3735
Db	1316	GlnProProGluTyrAlaIleThrAlaCysAspCysIleAsnArgSerGlnIleIleU--	1334
OY	3736	GTAACGATTAAGCAAAAAAAGCGGAAGAAAGCATATGCAATGTAAACACTGAT	3795
Db	1335	-----GlnProLysLysGluGluAsnValIleSerAlaCysLysIleValAlaIleU	1351
OY	3796	GATATACTTAAGAAACGATGGAAGAAACAAGTACGATGTGATGATCAAAAGAAAT	3855
Db	1352	LysIleLeuGluGlyLysAsnGlnIleArgThrIleValGlyIleCysAsnProLysGluSer	1371
OY	3856	AGTATGATATCCCGATTGGCAATGC--GGAATATTAATTTAGTGAAGACCTTGT	3912
Db	1372	-----TyrProAspTrpAspCysLysAsnAsnIleAspIleSerHisAsp--Gly	1387
OY	3913	GTCGTATGCCCTTAAAGACAAAGTTATGGCGATCTTGTGGCAATGATTAAGAA	3972
Db	1388	AlaGlyMetProProArgIleGlnLysIleCysLysLeuTyrTyrIleAlaHisGluSerGln	1407
OY	3973	ATAAAAAATTACATTCACAAAGTTAATTTAAAAAGACTTTCATCAAAATGTGCAGACGA	4032
Db	1408	ThrIleGlnIleLysThrAspAspAsnIleLysAspAlaPheIleLysThrAlaIleAla	1427
OY	4033	GAAACATTTCTTCATGAGTATTTATTTAAAGTAAG--GATGGTGAAGAAATGAATC	4089
Db	1428	GluThrPheLeuSerTrpGlnTyrTyrLysSerLysAsnAspSerGlnAlaLysIleLeu	1447
OY	4090	GATTAAGATTTAAAAAGAGCAAAATTCCTCCGATTTTTCAGATTCATGTTGTACAA	4149
Db	1448	AspArg-----GlyLeuIleProSerGlnPheLeuArgSerMetLeuThr	1463
OY	4150	TTTGGATATATAGAGATTTTATTTTGGAAACAGATATATACAAAAGTCATGTGAGGGA	4209
Db	1464	PheIleAspTyrIleArgAspIleCysLeuAsnTrpAspIleSerLysGlnAsnAspVal	1483
OY	4210	AGTAACATAAAGACAAATAGATTCTCTTTCAAAATGTGGACACAAATATCTCCAAAT	4269
Db	1484	AlaLysAlaLysAspLysIleGlyLysPhePheSerLysAspGlySerLysSerProSer	1503
OY	4270	GGAACAAACGCCACAGAAATGTGGACAGAACATTAATCATATGGAATGGCAAGCTATGTA	4329
Db	1504	GlyLeuSerArgGlnIleTyrTrpLysThrAsnGlyProGluIleTrpLysGluMetLeu	1523

OY	4330	TGCACTAGTAAATAATGGGCAAAAAAGATGAT-----TTTACGAAATACAC	4380
Db	1524	CysAlaLeuThrLysTyrValThrAspThrAspAsnLysArgLysIleLysAsnAspTyr	1533
OY	4381	GGTTACACAAACCGCAAAATTTAGTGCACAA---AGCACACACTTGGAGCAATTTGCCAA	4437
Db	1544	SerTyrAspLysValAsnGlnSerGlnAsnGlnLysProSerLeuGlnLysIleAla	1553
OY	4438	CGACCCAGTTTTCAGTATGGCTTACCGAATGGTATGACAGTATTTGCTATACAGCAA	4497
Db	1564	LysProGlnPheLeuArgTyrMetIleGluTyrGluGluPheCysAlaGluValGln	1583
OY	4498	AAATATTGGAAGATGTGCAGCAAAAATGTAACTCA--ATATGACCAATTGAATGTGAT	4554
Db	1584	LysLysGluAsnIleIleLysAspAlaCysAsnGluIleAsnSerThrGlnGlnCysAsn	1603
OY	4555	ACA-----GAATGTAATTAAGAAATGCGAGGACTACGTAATATATGAAA---AAA	4602
Db	1604	AspAlaLysHisArgCysAsnGlnAlaCysArgAlaTyrGlnGluIleValGluAsnLys	1623
OY	4603	AAAAAGATGGATTCACACAGATAAATATTTACACAGATGAACGGACAAAAAAGATTC	4662
Db	1624	LysLysGluPheSerGlyGlnThrAsnAsnPheValLeuLysAlaAsnValGlnProGln	1643
OY	4663	GATGACACACACATTTGGTGTATATGCTTACAGACTATTTACTGACAAATGCAACAGATTAC	4722
Db	1644	AspProGluTyrLysGluTyrGluTyrLysAsp-----	1654
OY	4723	TTGAACGAAATTTACTGCTACTGTGGTAAAGCCGAAAGTCCGTCTGTGGTACAA	4782
Db	1655	-----GlyValGlnPro-----IleGln	1660
OY	4783	AGAAATATACAAATTTAGAAAAACAGGCTTACTATGATCCGACAAACATTTGGGTGC	4842
Db	1661	GlyAsnGluTyrLeuGlnGlnLys-----CysAspAsnAsnLys--CysSerCys	1676
OY	4843	ACAAATTTATGAAATGTACAGCAAAATTTACTACATTTGCAAGTAAAGATTAAGTCAAA	4902
Db	1676	-----	1676
OY	4903	GGATTAGTAAAGGAGGCAAAACAGAGTGTCTAATTAAGTGCAAAACAAAGTCCATTAAC	4962
Db	1676	-----	1676
OY	4963	TACAAATTAAGTGAAGAAGATTTAGCTGAAGATGTGGCTTTTCTTCTGTGCACTACGATATA	5022
Db	1676	-----	1676
OY	5023	TGTTTCATGATTTGCATGGCAATTTTACAGATCCAGAACTTAAGATGAATAATGGGTGG	5082
Db	1677	-----MetAspCysAsnValLeuSerValSerProLysGluLys-----	1689
OY	5083	CGAAAAAGATTGAGGAAGTGGGGCGGAACGGAAGGATACAAATTTGGGTCAATTTACTACAAA	5142
Db	1680	-----ProPheGlyLysTyr-----	1684
OY	5143	GAATAAAAAAGAAAAAGAAATTAATAAAGCTGCGATCCGACAAATATTTCTTATAGCTGC	5202
Db	1695	-----AlaHisLysTyrPro-----	1699
OY	5203	CCGCTTGTATGTCTATGAATATTAAGTTTATGATTTAAGATTAATTAATTTCTAGATATT	5262
Db	1699	-----	1699
OY	5263	GATTAATTTGGAAGATGCAAAAAACAAAGACCGAGGAAATTTGACAAAAATTTTACAAA	5322
Db	1699	-----	1699
OY	5323	AATGGAACATCATTTGGCAAGAGAGATGATAGTACTACAGAAATCCGGTAGTACTCG	5382
Db	1699	-----	1699
OY	5383	CGAAAAATTTTCTGGAACGAAATAAAGAAATGTGTGTGCAACGCAATGATATGGCGGTAC	5442

Db	1699	-----	-----	-----	1699
Qy	5443	AAACGTGTAGCGATGATGAAATAGTGAATAGTGCAGAAGTGAAGATCTTAA	55020		
Db	1699	-----	-----	-----	1699
Qy	5503	AAATGTGTCTGACTTCAAGATGATGATTATCTATGCGGAAATTCGCGATGAAGT	55526		
Db	1699	-----	-----	-----	1699
Qy	5563	ACTGCGTACAGTTCTTCATGTTGTCGGAATGGGATGAAGATTTTGCACATATA	56222		
Db	1699	-----	-----	-----	1699
Qy	5633	GAAAGCAATTTGAGAAATTTGTAGGGCGCTGTAATGATTAATCTTGTCGTAAATGA	56822		
Db	1699	-----	-----	-----	1699
Qy	5683	GATTAAGAAAGAAATGTACAGATCGCTGACCAATATTAATAATTTATAGTACGG	57422		
Db	1700	-----	-----	-----	1702
Qy	5743	AAACACAGTATGAAAAACAAATCAAAATATGTGAGAAATAAGACAAATATATTC	58020		
Db	1702	-----	-----	-----	1702
Qy	5803	GAGCATCTGTGGCAAAAAGTCAGAGAGCGCTCGCAATATTAGACAAACAATTA	58622		
Db	1702	-----	-----	-----	1702
Qy	5863	AAATTTGTGAAATTAAGTGGAGATGTGATTAAGTATGAAAGATGTGCACA	59222		
Db	1703	-----	-----	-----	1704
Qy	5923	CAGCATTAAGTATGATGTAATAGTCAAAATATTCGCCGATCTAGACGATGAACCA	59822		
Db	1704	-----	-----	-----	1704
Qy	5983	GAAGTTGAAGCAAGTGTAAATTTGTCAGTCCGACGAGTCCACGCTACGAAGGGA	60422		
Db	1705	-----	-----	-----	1715
Qy	6043	ACACGTCACACCGGTATCACTGATATCAAAAGCAGCGCATCGAAAAAGACGAA	61020		
Db	1715	-----	-----	-----	1715
Qy	6103	ACAGCGCGCTACAAAAACGCGGAAAAAGTGAAATCTAACACAGAAATGCGACA	61622		
Db	1716	-----	-----	-----	1723
Qy	6163	CAAAACGCAACCCGACGACGACCAACAACAACGAAAAAGCAATCAACAGCACACA	62222		
Db	1724	-----	-----	-----	1728
Qy	6223	ACAGATCTGACGTGGCGCAACATGTTAAAGCCATCTTCGATTAACACAGATGCAAG	62822		
Db	1729	-----	-----	-----	1745
Qy	6283	GCTGGAATAGAGGTTGTAAATCCAAAACGATGACAAATAT-----CTTAATGGGGT	63362		
Db	1746	-----	-----	-----	1764
Qy	6337	TGTATT-----CTAGTACGCTAAACAAAT	63632		
Db	1765	-----	-----	-----	1784
Qy	6364	GAATAATGCAATATGATCCCTCTAGAGCAAAAAATTAATGATTAATTAATACAAAT	64232		
Db	1785	-----	-----	-----	1803
Qy	6424	TTAATTTATGAACCT-----GAAATTAAGCGTAC	64532		

OY	7171	GAGAAAGCAAAAGTCGTCTCATCAAAATTTTAAACATGAAATTTAAAGAAATTAATGTGGCAAT	7230
Db	2175	-----	2175
OY	7231	AAGGATTTGTTCTGTGATGCAAAAACCTTCTTCACAACTACCAAAACCAACAAATCA	7290
Db	2176	GlyPheCysGluThrLeu-----Asn	2177
OY	7291	CAATCATCCGATGCTAAATGATATGCGCAATGCTGATTAATGTCTCTGAGAAATTTAAC	7350
Db	2183	ThrSerSerLysAlaLysAspPheLeuLysThrLeuLys-----Thr	2188
OY	7351	AAGTGTAGCTGCCGAACCTTCAAAAAGGACATCTAGTT-----	7392
Db	2196	-----ProCysLysProAsnAsnValGluLysThrIlePheAspAspLysThr	2213
OY	7393	-----CATCAAAAAAAATTTACTGAACCT-----AAATACCTATGAAATGTGTAGAG	7440
Db	2214	PheLysHisThrLysAsp-----CysAspProCysLeuLysPheSerValAsnGly-----	2230
OY	7441	AAAGCAGCATATTAATTATCTAAAGAAAGCAAAATATATGATATTACTTGAAGGAA	7500
Db	2230	-----	2230
OY	7501	AAATTTATACCTAATGAGCTCACAAAGCAAAAGAAAGTAAATAGTTGACATAATAT	7560
Db	2231	-----LysLysAspGluCysAspAsnSer-----	2238
OY	7561	AATCCTTGAGATCCATAGAAACCTTATGCACCTGATTAATATATAGCAAGAAAGAACCTT	7620
Db	2239	-----LysGlyThrAsp	2242
OY	7621	TGTGAAATATAGAAAGAAATCGTTTAAAGTAGATTAATGATGAAATGTTACAAAAT	7680
Db	2243	CysArgAsn-----LysAsn	2247
OY	7681	TCAAGCTTCATCAGAGGAGAAAAAAGATATGTGTACCTCCAAAGAGCAACATATGTGC	7740
Db	2248	Ser-----	2248
OY	7741	TTAAGCAATTTAGATGAAATTTAAATTTGAAAGCTTAAAGATAGTAAATTAATCTCCATAAA	7800
Db	2249	-----IleAspAlaThrAspIleGluAsnGlyValAspSerThrValIleGluMet	2265
OY	7801	ATGTGCTGTGCACACTGCAGCAAAATGAAAGAAATAGCATTAATTAATAAACTTCAACTCA	7857
Db	2266	ArgValSerAlaAspSerLysSer-----GlyPheAsnGlyAsp	2278
OY	7858	-----GAGAACGGGTGCGCGCAATGATCCAAATGTGATGACTATGAAATTAATGATTTCGCT	7911
Db	2279	GlyLeuGluAsnLacys-----	2284
OY	7912	GATCTGGGTGACATAGTTAGAGAACACAGTATGTTACGAATTTGGTGTACTTACCTCCC	7971
Db	2285	-----ArgGlyAlaGly-----	2288
OY	7972	GTAGAAATAAATTTATATATAGGTTTGTGATACATATATGAAATATGAGAAATTAATAAT	8031
Db	2289	-----IlePheGluGlyIle-----	2293
OY	8032	AAAGGTGAAATTAATATACACAGATGTACAAACGTTTCGTTCTGCTGTGGAGTGCTAAT	8091
Db	2293	-----	2293
OY	8092	AGAAAAGATATTTGGAAAGCA-----ATGACGTGCAAAAGCACAGAA	8133
Db	2294	ArgLysAspGluThrLysCysArgAsnValLysGluTyrValValCysLys-----ProGlu	2312
OY	8134	GATGCAAAACTTTTAAAGAAAGAAAGATGATGATTTTAAGCATTAACATTAATATACAA	8193
Db	2313	AsnValAsnGluValLysGlyLys-----	2321

QY 8194 GATPAGTGTGACATGAGACGATCCAGCTTGATGATATATACCTCAACGTTTCA 8253
 Db 2322 -----HisIleIleGlnIleArgAla 2328
 QY 8254 TGGATGACGATGAGTGTGATATATATGTAAGCACTGATGAGAGATGGAATAATTT 8313
 Db 2329 LeuValIysArgTyrValGluTyrPhe-----PheGluAspTyrAsnIle 2344
 QY 8314 AAAAATCATGTGATCACTGTAAACATCTGACAGATGACAGATGATGATGATGATGAT 8373
 Db 2345 LysHis-----LysIleSerHisArgIleLysAsn-----Gly 2355
 QY 8374 AAGTGCACAGCTGTAAACGAGATGTCAGAGATATATAAATTTGTTCTTAATGGA 8433
 Db 2356 GluIleSerProCys-----IleLysAsnCysValGluIleTyrPhe--- 2368
 QY 8434 TCTCTATGATATACATGAAATTAATGAAAGATGATGATGATGATGATGATGATGAT 8493
 Db 2369 -----ValAspGlnIleValGluTyrPheGluIleThrGluArg----- 2382
 QY 8494 AAAATCTCTACTATGATCATGTTCAAAATTTGTACAAAAGTTGAAACCTTTTAAAGT 8553
 Db 2383 -----PheLysAsp 2385
 QY 8554 GAATGTTCTGTGAGACCTTTTCTGAATATCTTCATGAAACAGTAAGTGTGATATAT 8613
 Db 2386 Gln-----Tyr 2387
 QY 8614 AAATTTAATGAAATGATGGTCTTCCATATATACAGACATATGTTTCAAGAAACACA 8673
 Db 2388 Lys-----AsnAspAsnSerAspAspAsnValArgSerPheLeuGluThrIleLeuPro 2406
 QY 8674 AAA-----AGTTATAAAGAACT 8691
 Db 2407 GlnIleThrAspAlaAsnAlaIleValAsnLysValIleLysLeuSerLysPheGlyAsnSer 2426
 QY 8692 TGCAGTGTGACACTACCTCTGATAGATCCATGATGATATGTCTACCGATCAAAACAA 8751
 Db 2427 CysGlyLysSerAlaIle-----AsnGluGlnAsnLys 2438
 QY 8752 GATGATGATGAGAAATTAACAACCTTTACCTCTGCTCAGAAATGATATGATATATAT 8811
 Db 2439 AsnGly----- 2440
 QY 8812 CTGTGATATTTGGAACGATACCTGTTCTTAATAGTTGAGATGATTAACAAAGTGTATG 8871
 Db 2440 ----- 2440
 QY 8872 ATTCTCCAGAGAGAGACATTTATGACAGACCTATCACTGATATATATATGAAAA 8931
 Db 2440 ----- 2440
 QY 8932 GGTGATTAAGAAATTTTAAAAAACTTCTTACTCTGCTTCACTCAAGACAAATG 8991
 Db 2440 ----- 2440
 QY 8992 TTAGCTCAAAAAATTAATGCGAAGAGAGTGTGCTTTGAGCAATGAAATATAGTTAT 9051
 Db 2440 ----- 2440
 QY 9052 GCAGATTTATCCGATATATTAAGAACTGATGATGACACTTCATATATCTGAAAAA 9111
 Db 2441 -----GluTyrIleLysAspAlaIle-----AspCysMetLeuLysLysLeuLysAspLys 2456
 QY 9112 ATTAATAAATATTTGAACATCAATGAAAGCAACGAAATCTTAATAACATGCTGGGAA 9171
 Db 2457 Ile----- 2457
 QY 9172 AATATATGACCTCAGATATGCGACGCTATGTTATGATGATTAATAATGCTACTTCAAA 9231
 Db 2457 ----- 2457
 QY 9232 GTAACATTAGATGAGAGATGTGTCAATTAACAAAGATGAGAAAGTATACAGTTTCTT 9291

Db 2458 -----GlyLysCys----- 2460
 QY 9292 CGTTGTTAATGGAATGGCAACAGCATGTAGGAAAGAAACATGTAGTATCA 9351
 Db 2461 -----GluLysLysHisHisGlnThrSer 2468
 QY 9352 TTAATAACAAATGT-----CCTGTTCAACGAAGATTAATTTGAAAGCG 9396
 Db 2469 -----AspThrGluCysSerAspThrProGlnProGlnThrLeuGluAsp--- 2483
 QY 9397 TCAGATTTATTAAGACAACTCGATGTCAGATGATATTAATAAATATATAGCTTGAT 9456
 Db 2484 -----GluThrLeu-----AspAspAspIleGluThrGluAlaLysLys 2497
 QY 9457 ATATTGATTAATAAATACATGAAATCTAAATTAATTAATTAATTAATTAATTAATTA 9516
 Db 2498 AsnMetLeuProLysIleCysGluAsnVal-----LeuLysThrAlaGlnGlnLysAspGlu 2516
 QY 9517 TCTTCAGTATATATATGACATTAACATCTGAGAAAGATGTCAGTATATATTAATCA 9576
 Db 2517 -----GlyGlyCysValProAlaGluAsnSerGluProAla-----Ala 2530
 QY 9577 AAAGATCTCAATCGCTTGAGATTAAGATATTAATGAATATGATACAGAACAAA 9636
 Db 2531 ThrAspSer----- 2533
 QY 9637 AATATGAAATTAATGATTCACAAAGTACTAAATAAATTAATATCTGTTATATTTT 9696
 Db 2534 GlyLysGluThrProGluGlnThrProValLeuLys----- 2545
 QY 9697 GTTGAAGATGAACACACAAAAATCATGATGATGATGATGATGATGATGATGATGAT 9756
 Db 2546 -----ProGluGluGluAla 2550
 QY 9757 ACAGTGTGCTTAAGACATCTATTTCTTACACCCATGATGATGATGATGATGATGATG 9816
 Db 2551 ValProGluProPro-----ProProProProGluGluLysAlaProAla 2565
 QY 9817 CCTTATTTCTCAACATCAGTAGACACATATATATCTTAATAAT 9861
 Db 2566 Pro-----IleProGlnProGlnProProProProProProProProProPro 9861
 QY 9862 -----GATATATGAAAGTAGATGCTGTTGTTATGATGATGATGATGATGATG 9909
 Db 2580 LeuAspAsnProHisValIleuThrAlaLeuValThrSerThrLeuAlaTrpSerValGly 2599
 QY 9910 TTG-----ATAGCGCTCATTTCTGATGAAGAAATCAATCCTCTGTG--- 9954
 Db 2600 IleGlyPheAlaThrPheThrTyrPheThrLeuLysLysLysLysLysLysLysLys 2619
 QY 9955 GACTTGTGCGTATGATGATATCCCGCAAGAGATGATGATGATGATGATGATGATGATG 10014
 Db 2620 AsnLeuPheGlnIleLeuGlnIleProLysSerAspTyrAspIleProThrLysLeuSer 2639
 QY 10015 AAAAATAGTACATACCATATAGAGTGTCCATATAGGCAAAACATATATATATG 10074
 Db 2640 ProAsnArgTyrIleProTyrThrSerGlyLysTyrArgGlyLysArgTyrIleTyrLeu 2659
 QY 10075 GAAGAGATGAT-----AGTGAAGATGAGATTAATATATGAGATTTATCTCC 10125
 Db 2660 GluLysAspSerGlyThrAspSerGlyTyrThrAspHisTyr----- 2673
 QY 10126 TCTGATATCTCATCCGAAAGTGAATGATGATGATGATGATGATGATGATGATGATG 10185
 Db 2674 SerAspIleThrSerSerGluSerGluTyrGluGluMetAspIleAsnAspIleTyrVal 2693
 QY 10186 CCAGTGTGCTTAATATATAACATTTGATAGAGTATGATGATGATGATGATGATGATG 10245
 Db 2694 ProGlySerProLysTyrLysThrLeuIleGluValIleGluProSerGlyAsnAsn 2713
 QY 10246 ATACCAAGTAT----- 10257

[illegible]

Db 374 ProTtIleAspSngInGluNylsGluPheGluNylsGluSylsYrAlaGlu 393
 QY 1219 ATA-----CAATCATATTATCGAAGCAGTAAACAAATTTGCAAT 1257
 Db 394 IleLysAlaGluGluNylsGluThrSerIleThrIleGluAspLysThrIleAsn 413
 QY 1258 AATATTATATGCGAATATTAAACAAATTTAAAGAAACCTTAAGCAACCAATATGCA 1317
 Db 414 AsnLeu-----TyrValGlnGluPheThrGlnGlnLeu---GlnSerTyrLys 429
 QY 1318 ACTAATGACACTTTTAAATTTACTAAATGAAGAAAGATTTGTAAGA----- 1368
 Db 430 AspValGluAspPheLeuGlnLysLeuAsnGluGluThrIleGluSylsYrProThr 449
 QY 1369 ---GGATTACAGAGAAAGAAAGATATTACTTTTACTAAGAGTGTGATGCAAGGATA 1425
 Db 450 ValGluGluThrAlaAspProValAspPheThrLysGluSylsThrGluLys---Thr 468
 QY 1426 TTTTATGCTTCAGAAATTTCCCAAGTGTGCCGACTGCCGGCTCAAAATGTGATGTATA 1485
 Db 469 PheSerThrThrGluThrLysArgAlaCysProThrPcysGluLys----- 484
 QY 1486 AATATCACACAAATGCAATGATGATCGTGAAGCTGTAAT-----AATGAAGACTAT 1539
 Db 485 -----ArgLysAlaAspGlyThrTrpGluThrLeuAspMetAsnLysCysVal 501
 QY 1540 AAACCTTCATGGGTGTGAGCCCTACTAATATCACT-----GTCTTTATAGGT 1590
 Db 502 LysGluLysLysThrLysGluLysLysAsnIleThrAspIleProValIleThrProAsp 521
 QY 1591 AATGACACAGGTGATTTATACACAAATTTAGAAATTTTGTAAACAGCTCACTAATAC 1650
 Db 522 LysSerIleThrGluThrLeuAspLysLysAsnLysPheCysAsp---AspThrAsnGly 540
 QY 1651 AAGATTAATAATAATA-----AATGGGAATGCTATATAG----- 1689
 Db 541 LysAsnSerAsnGluNylsGluLysTrpGluCysTyrLysLysAsnGluLysAsp 560
 QY 1690 -----GATGAAATATATATATGATGTAACCTGAA-----CAAAATACT 1728
 Db 561 AspValAsnGluAspSerAsnIle-----CysValLeuGlnAsnAspLysGlnAsnThr 578
 QY 1729 GAAATCAATATGCTATCTTAATGATATATCATTTTCAATTTTGTGAATTAAGCGTT 1788
 Db 579 GluGluGluLysAsnAspArg-----SerTyrAsnSerPhePheThrLysTrpVal 594
 QY 1789 ACAATTTTAAAGCGATCTATTAAGTGAATGACAACTTAACCTTAACCTTGATTAATAT 1848
 Db 595 ThrGluThrLeuIleAspSerIleAspTrpArgLysGluLeuLysArgCysThrAsnAsn 614
 QY 1849 ACAACACGCAATGCTATGAT-----GAATGTAAACGAATTTGCTTATTTTGAACATG 1905
 Db 615 LysArgGluLysCysLysAsnLysLysCysAsnAsnAspCysLysCysPheGluArgTrp 634
 QY 1906 GTTAAACAAAAGAGAGAGATGGAATAGTAAAGAAAGCTGTCACAAAAAAAAGAAAT 1965
 Db 635 IleGluGluLysLysAsnGluTrpGluLysIleValAspTyrPheLysThrGlnProGly 654
 QY 1966 ATACAGCAATCGTATATAGTAAATATATATATCTTTTGAAGGTATTTTAAAGTT 2025
 Db 655 Phe-----AspIleLeuSerAsnAsnTyrAspLeuAlaLeuLysThrLeuLeuAspVal 672
 QY 2026 ATGATTAATCTGACAAAGATGAGCAAAATGGAAGAACT-----ATGGAATAATA 2079
 Db 673 GluAspIleLeuThrAsnIleGluAspGlyTyrLysGluValLysGluSerGlnHisIle 692
 QY 2080 AAAAGAAAAAAATGAGTTTCCAAATTTGGAATAATAGGCACTTTTAGAC----- 2133
 Db 693 ArgLysMet-----LeuGlnAsnGluLysAsnGlnGluGlnAlaAla 706
 QY 2134 -----AATGCAATAGAACTCTTGTAGATCACTTAAAGAACT 2172

Db 707 ValAlaValThrValLysArgAsnThrIleAspLeuLeuIleGluHisGluLeuAspGlu 726
 QY 2173 GCCACGATATGTAAGCAATTAAT---ACAACGACAGCATGT-----GAACATCC 2220
 Db 727 AlaAspLeuCysIleuAspThrHisIleGluLysProLysPcysSerAspGluAspAsp 746
 QY 2221 CATTAATGCA-----ACAACAAACCGCTGTCTTAACCTGTGGAGCCACCAACC 2271
 Db 747 HisGluGluGluValTyrValAsnAsnProCysAlaLysProSerGlySerTyrProSer 766
 QY 2272 ACTTAAATATTAAGAAATATGACCAATACTTTTAAAGAGTGCATGACGAAAGACGA 2331
 Db 767 LeuAlaAsn-----GluAlaIleHisGlnMetHisLysAlaIleLysThrGlnLeuSer 784
 QY 2332 AATCGTGCTCTCAATAA---TTGAAAGAAAGGACACACAGAGTATATTAACGTGG 2388
 Db 785 GlyArgGlyGluArgArgThrLeuLysAlaAspLysSerLysGluTyrLysArgGly 804
 QY 2389 GGTAGAGAAAGGCACTTCAGACAAATTTATGTGAATTAATGAATTAACATTTCAATCGT 2448
 Db 805 GlyGlnLysAsnThrLeuAsnGluAspIleCysLysIleThrLysAsnTyrThrAsn--- 823
 QY 2449 AATCTGCTTTTCAATGACCATGTGATGCGCAAGGACACAGGTGATGATCAACAA 2508
 Db 824 AspIleAspSerAsnGlyGlyAlaCysLysGlyLys-----AspGlySerAsnGlu 840
 QY 2509 AGATTGCTGTAGACATGATGGAAGTGAAGTGCAGACACATGCTTAAGATCAGAA 2568
 Db 841 ArgPheLysIleGluThrGluTrpSerAsnValGluAspLysLysThrIleSerTyrSer 860
 QY 2569 GATCTTATTTGCTCTCTAGACAGCAATATATGATATCAATTTTGAACATTTACAA 2628
 Db 861 AspValTyrLeuProProArgArgGluHisMetCysThrSerAsnLeuGluAsnLeuAsp 880
 QY 2629 ACGGATGATCACCCACTTAATGGAATATGTTGATGATTAATTAATTCCTTTTG 2688
 Db 881 ValGluSerValThrGluAsnGlyLysAlaIle-----HisSerLeuLeu 895
 QY 2689 GGGGATGTTCTTATGACGAAATATGAAACAAAGATTAATGCAATGATTAAGAA 2748
 Db 896 GlyAspValIleLeuThrAlaIleThrAlaMetAspAlaIleGluIleAsnArgTyrLysSer 915
 QY 2749 AAGAAATACCTTAAGGCCCCCAAGAAAGTAACACCAACCAACGACACTATCTGT 2808
 Db 916 GlnAsnSerIleGluAspPro-----IleAspLysLysHisGlnGluSerIleCys 932
 QY 2809 CGACCTATAGCTTACATTTTTCAGATATAGGTATGATTAATTCGAGAGAGATCTGTG 2868
 Db 933 ArgAlaIleLeuArgTyrSerPheAlaAspIleGlyAspIleLeuArgGlyAspIleTrp 952
 QY 2869 GAAAGAAACGGTGCATGCTAAGCTCAGAGCAAGGACATTTGGAACTTTTGGTAAATA 2928
 Db 953 AspAsnGluThrGluMetAsnHisIleLeuLysGlyHisIleGluLysValPheLysAsnIle 972
 QY 2929 CATATGCTACTCAAGGCAAGAAATGATATATATGATGCTGCC-----CCC 2979
 Db 973 LysGluLysLeuProGlyIle---GlnAspLysTyrThrGluAspGluLysAsnAspPro 991
 QY 2980 AAATATTTAAATGAGGAGAAATTTGGGGAACCTATAGAGCAAAATGATGGAAGCC 3039
 Db 992 ProTyrLysLeuLeuArgGluAspTrpTrpGluAlaAsnArgHisGlnValTrpArgAla 1011
 QY 3040 ATGAATGTGATATTAATATTTGAAGATTAATCGGACACCAATCAACAAAGATAGT 3099
 Db 1012 MetLysCysAlaThrLys-----LysAspLys-----AsnMet 1022
 QY 3100 TATTCGCGATATGATGATACACCATTTGATATATATCCCAAAATTAAGATGG 3159
 Db 1023 LysCys-----AsnGlyIleProIleGluAspTyrIleProGlnArgLeuArgTrp 1039
 QY 3160 ATGACCAATGCGCAGATGCTACTGCAAGGTCAGAAAAAGAGATATGATTAAGTAAAG 3219
 Db 1040 MethThrGluTrpSerGluTrpPheCysLysMetGlnSerValGluTyrGluLysLeuLys 1059

[illegible]

Db 1507 ----- 1507
QY 5356 ACTACAGAAATCCCGTAGTACTGCGGCGAAAAATTTTCTGGAACGAAAAATAGAAATGT 5415
Db 1507 ----- 1507
QY 5416 GTGTGGAACGCAATGATATGTCGGGTACAAACGCTGATGAGATGTAATAATAGTGAAT 5475
Db 1507 ----- 1507
QY 5476 AGTCAAGAAGTGAAGTGAATGTAATAATAATGCTTCTGTAAGTATGATATAT 5535
Db 1507 ----- 1507
QY 5536 CCTATGGGAAAAATCGCATGAGAGTACTGCGTATCACTTCTGATGCTTGGCGGAA 5595
Db 1507 ----- 1507
QY 5596 TGGGGTGAAGATTTTTCGAAACATTAAGAAAAAGAAATTTGAGAAATTTGGAGCGCTGT 5655
Db 1507 ----- 1507
QY 5656 AATGATTACTTGTGTGATATGAGATTAAGAAAGAAATGACAGATGCGGTACA 5715
Db 1508 AAspBglYserCysHisLysAsnGlyGluGluCysGlnLysCysArgGluGluCysLys 1527
QY 5716 CAATATAAAAATTTTATGATGAGTGAACACACATGATGAAACAAATCAAAAAATAT 5775
Db 1528 LysTyrGlnLysTyrLeuGlnLysTyrPargLysAsnTyrAspLysGlnLysValLysPhe 1547
QY 5776 GGTGGAATTAAGACAAAAATATATCCGAGCATCTGTGGCAAAACATGCAGAGACCT 5835
Db 1548 LysThrAspLysGluLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1566
QY 5836 CGCGAATATTTTGAACAACATTAATAAAAAATTTTGTGAAAAATAAGTGA 5886
Db 1567 TyrGluTyrLeuGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1586
QY 5887 GATTGGAATATATGATGATGAAAGATGTGTCCACACAGCATTAAGTGGTAAAT 5943
Db 1587 AspCysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1604
QY 5944 AGTCAAAATATATGCGCCCATCATTAAGACGATGAACCAAAAGAAATGGAAGAAATGTAAT 6003
Db 1605 ThrAspAsnMetProGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 1624
QY 6004 TGTCAAGTGGCAGAGTGCACACGCTGATGAGAAAGGAAACACGCTCACCACGCGTATCA 6063
Db 1625 CysAspSerProThrLysLysProGlnValLeuProValSerPro 1639
QY 6064 CTGATATCAAAAGCGAGCGCATCGAAAAAGAAAGCAAGAAACAGCGCGCTACAAAAACAG 6123
Db 1639 ----- 1639
QY 6124 CCGAAAAAAGTGAATAATCTAACACAGAAATGCGAGCACAAACAGAACCCGACGACGA 6183
Db 1639 ----- 1639
QY 6184 GCACAAACAACAAGAAAGCAACATCAACAGCAACAACAAGAAATCTGACGTGGCACA 6243
Db 1640 ----- 1647
QY 6244 ATGGTAAAGCCATTCTTTCGAAATAAACAGATAGAGGGGTGGAATAGAGGGTTGTAAT 6303
Db 1648 IleValAspIleLeuAspProAsnSerGlyThrGlyLysAsnAspAlaCys 1665
QY 6304 CCAAAACGATGGACAA-----TATCCTAAATGGGTGATATAGAGTAAG 6351
Db 1666 GlnGluLysTyrGlnLysGlnLysLysLysLysLysLysLysLysLysLysLysLys 1685
QY 6352 TCTAAA-----GAAATGAAATGGCAATATGATGCTCTGAGGAAAAAATATATGT 6405

Db 1686 IleGluLysGlyLysGluAspAspValValCysIleProProArgArgGlnLysLysLys 1705
QY 6406 ATAAATTAATATACATATTTAAATTAATGAACCTGAAATTAAGCGTACATGATATATAA 6465
Db 1706 IleLysGlnLeuLys-----GluPheSerGlyLysLysAsnAspGlnLeuArg 1721
QY 6466 GAGCCTTTATTAATAGTGCAGCATATGAAACCTCAATTTTGTGCTTAAATATATAT 6525
Db 1722 LysAlaPheIleGluCysAlaAlaValGluThrPhePheAlaThrPheLysTyrLysGlu 1741
QY 6526 GAA----- 6528
Db 1742 AspLysLysAsnGlnLysLysGluGluThrSerTyrLeuValArgLeuGlyPheLeu 1761
QY 6529 -----AATCCTGACAGAAAAATGCAAAATGCAAAATGCAAAATG 6567
Db 1762 GlyLysGlnLysSerProAspGlnGluGlnLysAspLeuLysSerGlyLysIle 1781
QY 6568 CCGATGATTTTAAAGATATGATATATATGATGATGATGATGATGATGATGATGATGAT 6627
Db 1782 ProAspGlnPheLysArgLysMetPheTyrPheGlyAspTyrArgAspIlePhePhe 1801
QY 6628 GGAACGATATTTCTAATGATTAATAAATAATTAATGTAACAATAGTGAACACCAT 6687
Db 1802 GlyLysAsnMetCylLysAspMetLys-----ThrValGlnLysLysLysSerIle 1819
QY 6688 CTCATGAAAAATTAATAAGAAAAACAGATTAATAAATAAAGATGAAATTCGTAAATA 6747
Db 1820 Phe-----ProAsnSerValLysProAspGlnThrArgGluThr 1833
QY 6748 TTTTGGAGAAAAATTAATAATTTTGGGAAGGATGATATATGATTAATCTTATCAT 6807
Db 1834 TrpTyrAsnAsnAsnAlaLysAspIleTyrPasnGlyMetLeuCysAlaLeuSerTyr 1852
QY 6808 CTCACAGCAGAAAAAGAAAAAATAAGATTAATATAC-----CAGTAC 6855
Db 1853 -----AspThrLysLysAspArgLysLysLysLysLysLysLysLysLysLysLys 1870
QY 6856 AATGACATGACCAACTGACG-----CCTTCCTGGAAGATTGTAAAA 6900
Db 1871 AsnAspTyrAsnLysValThrPheAspAsnGlyAsnThrLeuGlnLysIleTyrAlaThr 1890
QY 6901 AGGCCCAATTTTGAATGATGCTGCACAGAAATGGCGAGAAATTTGTATTAAGAGAAAG 6960
Db 1891 ValProGlnPheIleLeuSerPheGlnGluTyrAlaGlnLysPheCysArgLysAspGly 1910
QY 6961 GAACAGTTGTTAAATTTGAGCGCGGCTGTAG-----GAATATGAGTGT 7005
Db 1911 HisLysLeuAlaHisIleLysAspAspCysArgSerAspThrProGlyLysMetTyrCys 1930
QY 7006 AATGATGATATGACGGTAAAGACAAAGAA----- 7035
Db 1931 SerGlyAspGlyHisAspCysThrAsnGlnGluLysHisAsnAspMetPheAlaAsp 1950
QY 7036 -----TGTGACAGCGCGGTGTGAACATATCAAAATTTTATTAAGAAATG 7080
Db 1951 SerTyrCysProAspCysLysLysAlaCysArgLysTyrAsnLysTyrIleGluLysLys 1970
QY 7081 AAAATGATATGAAAGCAAGAAAGAAAGAAAGTCAAAAGATTA----- 7125
Db 1971 ValGlnGluPheTyrAsnGlnLysSerLysLysLysLysLysLysLysLysLysLysLys 1990
QY 7126 -----GATGCAAAAAGTAAAGATTTCTTCTACGTAAGAAAGACATAGAAAGCA 7179
Db 1991 AsnLeuAspAsnLysTyrValLysGlnPheTyrAlaThrSerGlnGly-----LysTyr 2008
QY 7180 ACATGTCATGATGATATTAACATGAATTAAGAAATTAATGTCATATAGATATGT 7239
Db 2009 LysSerValAspSerPheLeuAsp-----LysLeuLysGln----- 2020
QY 7240 TCTTGTATGCAAAAACCTTCTCACAACATCACAAAAAACAACACACATCATCATCC 7299
Db 2020 ----- 2020

```
OY 7300 GATGCTATGATATGCCAGATCCGTGGATATGTCCTGAGAAATTAAACAGTGTGAC 7359
Db 2021 -----ArgSerHis 2023
OY 7360 TGTCCGTAACCTTCAAAAAAGGATCTATGATTCATCAAAAAAATTACTGAACTTAA 7419
Db 2024 Cys----- 2024
OY 7420 ATACCTATGATTTGTGTAGAAAGACAGATATTATTATCTAAAGAGAGAAATTAAT 7479
Db 2025 -----His 2025
OY 7480 ATGATATTACCTTGAAGAAATTTATACCTATGAGCTCAAGAAAGAAAGT 7539
Db 2026 MetAsp---ThrLeuGluGlyLysIle---AspPhe 2035
OY 7540 AAAAATGTGGATATATATATCCCTGGATCCCTAAGAACCTTATGCACCTGATAAA 7599
Db 2036 LysAsn-----ProLeuLysThrPheSerSerSerThr 2046
OY 7600 TATATAGAGAGAGAAACCT-----TGCAAAATAGAGAGAAATTCGT 7644
Db 2047 TyrCys---LysThrCysProLeuTyrGlyValGlnCysArgAsnThrSerSphis--- 2064
OY 7645 TTTAAGTAGATATATGAATGGAATGTTCACAAAATTCAAAAGTTCTATCAGAGAGAAAA 7704
Db 2065 -----CysIleGlnAsnSergly---AsnGluLysLys 2074
OY 7705 AGAGTATGTCTACCTCCAGAGAGAACATATGTCCTTAAGAAATTAAGATTAATAA 7764
Db 2075 -----TrpGluHisAla---LeuAspThrIleLys 2083
OY 7765 ATTGAAGACTTAAAGATAGTATTAATTCCTCAAAAATGTCGTGCACTGACAGCAAT 7824
Db 2084 IleLysAsnGlyAlaProThrSerIleAsnValGlnMetIleAspArgy----- 2099
OY 7825 GAAGAAATGACATATAATAAAACTCAACCTCAGAGAGAGGGTGGCAATGAATCAATA 7884
Db 2099 ----- 2099
OY 7885 TGTGATACATATAAATATAGTTCCGTGATGTCGGTGACATAGTAGAGAGACATATAG 7944
Db 2100 -----ArgGly-----GlnTyr 2101
OY 7945 TTACGAATGTGTGTACTTACCTCCGAGAAATTAATTAATTAAGGTTTGAATAC 8004
Db 2102 -----GlnTyr 2103
OY 8005 AATATGGAATAAGAGAAATAAATAAAGGTAGAAATAAATACAGAGATGACAAACG 8064
Db 2104 Ile-----GlnIleHisSerGluAsnSerPheLysGlnSerArgLeu 2117
OY 8065 TTTCCTTCTGCTGTGGGATGCTATATGAAAAGATATTGG----- 8106
Db 2118 LeuLysSerVal-----ArgGluGlnLysTrpGluCysSerPheValAsn 2122
OY 8107 ---AAAGCAATGACGTGCAAAAGACACAGAAAGTGAACCTTTAGAAAAGAGAAATG 8163
Db 2133 LysLysMetAspValCysLys-----LeuLysAsnAspPhe---LysGluAsnIle 2147
OY 8164 GATGATTTGAACGCATACAC-----TTAATACAGAAATAGTGTGACATTAAGAC 8214
Db 2148 AspThrAspGluThrIleThrPheLysValLeuLeuGluAsn----- 2161
OY 8215 GATCACCTGTTGATGATATATACCTCAACGCTTTGATGATGACTGAATGGTCTGAA 8274
Db 2162 -----TrpLeuGlnAspPheIleGlu 2168
OY 8275 ---TATTATTGTAAGACAGTGAAGAAATTGAAAAATTAAATATATGATGACAC 8331
Db 2169 GlyTyrTyrIleSer-----LysArgLysIleAspIle 2179
```

```
OY 8332 TGTAAAAACATCTGCACAGATGCAAGAAATGATATGATGAAAAATAGGTGACAGTGA 8391
Db 2180 CysThrLysLysGlu-----GlnHisThrAlaIleGlnLysLys 2193
OY 8392 ACCAGATGTCAGAAATATAAAAATTTGTCTTAAATGGAATCTCTATTCGATATACAA 8451
Db 2194 SerLysCysGlu-----CysIleGlyLysIle----- 2202
OY 8452 TCMAATTAATACAAAGAAATGTATGACAAACCAATATATACAAAATCTCTATGAT 8511
Db 2203 -----LeuLysGlnLysThrThrGluTrpAsp 2211
OY 8512 CAGTTCAAAATTTGTACAAAAGTTGAAAACCTTTTAAAAGTGAATGTCTGTGAGAGC 8571
Db 2212 GluIleLysThrHisPheAsnLysGln----- 2220
OY 8572 TTTTGTGATATCTTCATGCAACAGTAAGTGTGATTAATTAATGAAATGAT 8631
Db 2221 -----AsnArgGlyAsp 2224
OY 8632 GGTTCCTCCAAATATAGCAACATATGCTTTCGAGAAACCCAAAAGTTATTAAGAGCT 8691
Db 2225 Gly-----TyrGluIleAlaHisLysValArgAsnTyrPheGlu--- 2237
OY 8692 TGCAGTTGACACTACCTTCAAGAAATCCATTTGATATGTCCTACCGATCAAAACAAA 8751
Db 2238 -----LysAsnAlaVal----- 2241
OY 8752 GATGATGTAAAGAAATTAACAACTTTTACCTTGTGTCGAGAAATGATATGATATAT 8811
Db 2242 -----Gln 2242
OY 8812 CTGATATATGGAACGATACCTTGTCTTATATAGTTCAGATGATTAACAAGGTATG 8871
Db 2243 LeuLysLysTrp-----IleAspAspLeuLysHisVal--- 2253
OY 8872 ATTCCTCCAAAGAGACATTTATGTACAAAGCACTATCACTGCAATTAATTAAGAAA 8931
Db 2254 -----LysLysIleAspAspLeuArgIleValLeuTyrAspCysAlaAsn 2269
OY 8932 GGTGTAAGAAATTTTAAAAAAACTTCTACTTCTGCTTCACTGACAGACAAATG 8991
Db 2270 MetAspLysGluIleLysProPheGluAsnGlnProHisGluThrThrSerProAsnTyr 2305
OY 9052 GCAGATTAATCCGATATATTAATTAAGAGACATGATATGACACTTCATTAATCGAAAA 9111
Db 2306 CysAspIleSerProThrHisIleLeuThrGluProSerGlnThrAspThrGluThr 2325
OY 9112 AATTAATAAATAATTTTGAACATCAAAATGAAGCAACGAAAATGTAAACATGCTGGAA 9171
Db 2326 LeuAsnProTyrAspGluThr----- 2332
OY 9172 AATTAATAGAGATGATATGACAGCTATGTATGTGATATTAATTTGCTACTTCAAAA 9231
Db 2332 ----- 2332
OY 9232 GTAAACATTAAGATAGAGATGCTCAATTAACAAAGATGAAGAACTAATAGTTTCTT 9291
Db 2333 -----ProGluAspThrSer----- 2338
OY 9292 CGTTGGTAAATGAATGGCAAGCAAGCATGTAAGAGAAAAGAAACATTAAGTATCA 9351
Db 2339 -----ThrSerSer 2341
OY 9352 TTTAAAAACAATAATGCTCTGTTCAAAAGAGATTAATTTGAAGCGTCAGAAATATTAA 9411
Db 2342 ArgProAsnPheCysPro----- 2347
OY 9412 CAACCTGATGTCAGAAATGATATTGAATAATATTAGCTTGAATATATGATTAATAAAT 9471
```

```

Db 2347 ----- 2347
OY 9472 ACATGGAATCTAATAATATAAGCAATTAAGATCAATCTCAGTAATATA 9531
Db 2348 -----GlnValGluProProProLysProGluVal 2357
OY 9532 GACATTAACATCTGAGAAAGTTCAGTCATATATAATCAAAAGATTCATAGC 9591
Db 2358 ProGlnLysPro---GluGluThrAlaGluAspThrGluAspThrGluGluAla 2376
OY 9592 GCTTGAGATTAAATATATATGAATAGTACAGAAACAAATATGAATATAT 9651
Db 2377 AlaAlaPro-----ProValAlaProSerSerSerGluGluGlu 2389
OY 9652 GAATTCAGAGAGTACTAAAAATATATCCGTTTATATTTGTGAAGATGAACA 9711
Db 2390 AlaProLysGluValAlaProGluLysLysPro----- 2400
OY 9712 CACAAATATCATGTACTAGATGAATATATAAGAAAGAAAGCAAGTCTCTTAA 9771
Db 2401 -----LysGluValProLysProGluProLys 2409
OY 9772 GCACGTATTTCTTACACCCATGTAGATTTCTTATCAAGACCTTATCTCAAG 9831
Db 2410 Ala-----ProLysLysArgArgProArgGluValThr 2420
OY 9832 CATCGAGTAGACAAATATGATCTTAAATAATGATATATGTAATCTCTGTT--- 9888
Db 2421 His-----SerLeuProGluMetValSerIleSer 2431
OY 9889 -----GTTATGTATATGCGCGTTAGGTTGATAGCGCTCATTTATGAAGAA 9936
Db 2432 AlaPheProLeuSerValIleAlaPheAlaLeuSerTyrPheValLeuLys 2451
OY 9937 AATTAATGATGCTGTGAGTCTGTGATATCGAATATATCCGGAAGAGTATGA 9996
Db 2452 LysThrLysSerThrIleAspLeuArgValIleAsnIleProLysGluPyrGly 2471
OY 9997 ATGCTACCTGTGATATCAAAATATAGTACATATATGAAGTGTCCATTAAGGC 10056
Db 2472 IleProThrMetLysSerLysAsnArgTyrIleProTyrAlaSerAspArgTyrLysGly 2491
OY 10057 AAACATATATATATATGAGAGAGTACTAGTGAATGAGATGAATATATATGAGAC 10116
Db 2492 LysThrLysIleTyrMetGluGluAspGluSer---AspAspTyrThrTyrIleGlyAsp 2510
OY 10117 TATCTCTCTGATATATCT---TCATCGAAAGTGAAGATGAAGATTTG 10164
Db 2511 IleSerSerAspIleThrSerSerGluSerGluTyrGluMet 2527

RESULT 9
ID 025733 PRELIMINARY; PRT: 2924 AA.
AC 025733:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE PfEMP1 variant 1 of strain MC.
GN MCVAR-1 PfEMP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MALAYAN CAMP;
RX MEDLINE=95530812; PubMed=7541722;
RA Baruch D.I., Pasloske B.L., Singh H.B., Bi X., Ma X.C., Feldman M.,
  Tarsach T.F., Howard R.J.;
  *Cloning the P. falciparum gene encoding PfEMP1, a malarial variant
  antigen and adherence receptor on the surface of parasitized human
  erythrocytes.*
RL Cell 82:77-87(1995).

```

```

DR EMBL: U27338; AAB60251.1; -.
DR InterPro: IPR000345; Cytic_heme_bind.
DR InterPro: IPR004258; PfEMP.
DR Pfam: PF03011; PfEMP. 2.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
FT VARIANT 104 104 S->T.
FT VARIANT 105 105 E->K.
FT VARIANT 139 139 D->G.
FT VARIANT 362 362 K->E.
FT VARIANT 503 503 K->N.
FT VARIANT 581 581 S->T.
FT VARIANT 593 593 M->V.
FT VARIANT 617 617 I->K.
FT VARIANT 631 631 V->G.
FT VARIANT 658 658 D->E.
FT VARIANT 722 722 K->N.
SQ SEQUENCE 2924 AA; 335860 MW; 7077553BDAC82B26 CRC64;

Alignment Scores:
Pred. No.: 3,95e-151 Length: 2924
Score: 3154.00 Matches: 1013
Percent Similarity: 38.74% Conservative: 453
Best Local Similarity: 26.77% Mismatches: 963
Query Match: 16.21% Indels: 1356
DB: Gaps: 150

US-10-087-013-1 (1-10628) x 025733 (1-2924)
OY 97 AAAGTGCAGAAATGTTTGAAGCTTATGCCAAATATA----- 138
Db 13 LysAspAlaLysHisAlaLeuAspArgIleGlyGluValTyrLysGluLysValGlu 32
OY 139 AGACATCATCAAAATATGCAAAAGAAATGTCATTCGTTGAAAGGATTTGACGAA 198
Db 33 AsnAspAlaGluLysTyrLysLys-----AlaLeuLysGluLysLeuGluGlu 48
OY 199 GCAGAAATTCGTGGTGCTCTTACGCCAGTAATATAGCATATATATATCAATAT 258
Db 49 AlaLysGlyIleGlyLeuAlaLeuAlaSer-----ProAsn 60
OY 259 CCATGTAATTTAGATCATTAAGAAACATTAAT---TTACGGTATGATGATGTAATTG 315
Db 61 ProCysLysLeuValGluAspTyrTyrAsnAspArgLeuLys-----ArgLys 76
OY 316 AGACATCTCTCCATGATAGAGAAACAAACCGATTTGATGATGAAGATCTGAATCT 375
Db 77 ArgTyrProCysAlaAsnArgIleThrValArgPheSerAspLysTyrGlyGluGlnCys 96
OY 376 GGA---ATTAATATACGTAATTTTAAAGAAATATGATCTATA---GCCTGTGGCCA 429
Db 97 ThrPheAsnArgIleLysAspSerGluAsnAspAsnSerIleGlyAlaCysAlaPro 116
OY 430 CCTAGAGACGACATATATGTTGATTAATAAACTTGAAAGCTTAATATGATTAATACCCA 489
Db 117 TyrArgArgLeuHisLeuLysCysAspTyrAsnLeuGluLysMetLysThrSerThr 136
OY 490 AATATTCATGATTTTGGAAATGTTAGTACAGCAAAATACGAAGGTGAATCAAT 549
Db 137 Lys---HisAspLeuLeuAspValCysMetAlaAlaLysTyrGluGluLysSerIle 155
OY 550 GTTAATATATAT---CCACATAAGAAC---TCAGACGTTGACT 551
Db 156 LysThrHisTyrThrLysHisGluLeuThrAsnProAspThrLysSerGluLeuLysThr 175
OY 592 GCTCTGCACAGATTTTGCAGATATAGGATATGTAAGAGATAGTATGTT--- 648
Db 176 IleLeuAlaLysSerPheAlaAspIleGlyAspIleAlaLysAspLeuTyrLeu 195
OY 649 -----AAACCAATGTCCATGACAAAGTGAACCGGCTCCGAGAGGTT 693
Db 196 GlyTyrAspAspLysGluLysAspGluLysLysLeuLysLeuLysAsnLeuIleGluIle 215
OY 694 TTCAGAAATATACATGATGATGAGATGAAGATGAATGAATTAATCAATCTGATGA 753

```

Db 216 PheylsLysIleHisGluAsnLeuGlyThrcGlnAspAlaLysAspHisLysLysAsp 235
Qy 754 TCTGGAATTTATTAATTAAGAGCAATGTCGAATGGAATTAATAATTAAGTTCG 813
Db 236 GluLysAsnLysLysGlnLeuAlaGluAspTrpThrAlaAsnAspSerThrValTrp 255
Qy 814 GAACCTATTAACATGTGATGCAATCATATTAATTCGATATTTTATGCAATGAGAAATAT 873
Db 256 LysAlaIleThrCysHisAlaGlyGluSerAspLysLysThrGlySerCysSer 275
Qy 874 ACACCATTTATTTCAATCTCT--AAATGCCCATTAACAGAGAA-- 918
Db 276 GlyLysLysLysAspLysCysArgCysLysAspGluLysAsnGluThrAsn 295
Qy 919 --GTTCCATCAATTTAGATTTGTCCTCAATATTTAGCTGTTGTCGAGCAATGGA 975
Db 296 GluValProThrLysPheAspLysValProGlnLysLysLysLysLysLysLysLys 315
Qy 976 GAAGAGTTTGGCCGAAAAAATAATTAATTAAGGAGCAAGCACTGCTGCTGAAT 1035
Db 316 GluAspPheCysArgLysLysLysLysLysLysLysLysLysLysLysLysLys 335
Qy 1036 GACAAA-----GACGCTATATTGATGATGATGATGATGATGATGATGATGAT 1089
Db 336 GluLysGlyLysAsnLysLysLysLysLysLysLysLysLysLysLysLysLys 1149
Qy 1090 TGGAAAAAGATTTTGGATTTGATTAATTAAGTCTGATGATGATGATGATGATGAT 1149
Db 335 ArgGlyAlaGluLysLysLysLysLysLysLysLysLysLysLysLysLysLys 374
Qy 1150 GTTTTGAAGTTTGGTGAAGATCAACAAAGCAATTTAAACAAACAAACAAATAT 1209
Db 375 ArgPheValLysTrpLysLysLysLysLysLysLysLysLysLysLysLysLys 394
Qy 1210 GAAAAAGAAATACA-----TCATATTATGCAAGATTAATTTTCAAT 1257
Db 395 AspGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 414
Qy 1258 AATTAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1317
Db 415 AsnLeu-----LysValGlyHisPheLysLysLysLysLysLysLysLysLys 430
Qy 1318 ACTAATGACACTTTTAAATTTACTAATTAATTAATTAATTAATTAATTAATTAAT 1371
Db 431 ThrValAspLysSerLeuGlnLysLysLysLysLysLysLysLysLysLysLys 450
Qy 1372 TTACAGAGGAAAGATATTTACTTTTACTAAGCTGCTGATGACAAAGGATTTTAT 1431
Db 451 ValGlyAsnGluLysLysLysLysLysLysLysLysLysLysLysLysLysLys 470
Qy 1432 CATTGAGATATTTGCAAGTGTGCTCCGACTGCGGCTCAATGATGAT----- 1479
Db 471 HisThrThrLysCysGluAlaCysProTrpLysGlyAlaGlnLysGluLysAsnGly 490
Qy 1480 GGTATTAATTAATCAACAAATCA-----GATATGATGCTGAGCTGATTAATGAAG 1536
Db 491 GlyTrpLysAlaLysGluLysSerCysAlaLysLysLysLysLysLysLysLysLys 510
Qy 1537 TATTAACCTCCATGGGTGTGAGCTATATCACTATCACTGCTTTATAGTGATGA 1586
Db 511 -----AsnSerThrAspLysLysLysLysLysLysLysLysLysLysLys 523
Qy 1597 CAAGCTGATTAATCAACAAATTAAGAAATTTTGTAAAGCTCACTAATTAATGAAG 1656
Db 524 ArgSerLysThrLeuGlnLysLysLysLysLysLysLysLysLysLysLysLys 542
Qy 1657 AAAATATATCAAAATGGAATGCTATTAATTAAGATGAAATTAATTAATGAATGA 1716
Db 543 -----AsnAspLysLysLysLysLysLysLysLysLysLysLysLysLys 555
Qy 1717 GAACAAATTAATCAATTAATGAT----- 1743

Db 556 AspAspGluThrAspAspSerAsnAspCysValLeuGlyAspTrpGlyAsnLeuThrLys 575
Qy 1744 AATCCAGATATATATATCATTTATTTTGAATTAATGCTGATCAATTTATTAAG 1803
Db 576 GluAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 595
Qy 1804 GATACATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860
Db 596 AspSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 615
Qy 1861 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
Db 616 CysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 635
Qy 1921 GAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
Db 636 ThrGluTrpGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 655
Qy 1981 TATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2040
Db 655 ----- 655
Qy 2041 AAAGATGAAGCAAAATGAAGAACTTATGAAATATTAAGAAAGAAAGAAAGAT 2100
Db 655 ----- 655
Qy 2101 TCCAAATTTGAAATTAATAGGACTATTTAGAGATGCAATTA-----GAACCTTTGTA 2154
Db 656 -----ThrHisAspAspPheLeuGlnLysLysLysLysLysLysLysLysLys 671
Qy 2155 GATCACTTAATTAAGAACTGCTGATGATGATGATGATGATGATGATGATGATGAT 2214
Db 672 GluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 683
Qy 2215 ACATCCCATTAATGCAACAAACAAACCCGTGTGTTAAACCTGAGAGCAGCAAC 2274
Db 683 ----- 683
Qy 2275 AAAATATTAAGAAATTAAGCAATTAATTAAGGATGATGATGATGATGATGATGAT 2334
Db 684 -----LysAsnGlyLysLysLysLysLysLysLysLysLysLysLysLys 694
Qy 2335 CGTGTCTTCAATTAATTAAGAAAGAAAGCAACAGAGATTAATTAAGTGGGTAG 2394
Db 695 -----GlyValAlaGlyLysLysLysLysLysLysLysLysLysLysLysLys 713
Qy 2395 AGAAGGACTTCAAGGACATTTATGATGATGATGATGATGATGATGATGATGATGAT 2454
Db 714 AlaGlu-----LysAspThrThrLysLysLysLysLysLysLysLysLysLys 729
Qy 2455 GGTTTTCAATGACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2514
Db 729 ----- 729
Qy 2515 GTGTGAGAACTGAATGGAAGTGAATCCGGAACACATGCTGAAGATCAGAGATGTT 2574
Db 730 ----- 730
Qy 2575 ATTATGCTCTAGAGAGACATATATATGATGATGATGATGATGATGATGATGATGAT 2634
Db 742 ThrCysProGln-----Glu 747
Qy 2635 GATCAACCACTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2694
Db 748 AspArgSerVal----- 751
Qy 2695 GTTCTTCTATCAAGCAATATGAGCAAAAGATTAATTAATTAATTAATTAATTAAT 2754
Db 752 -----AlaArgSerGluSerAlaThrVal----- 759
Qy 2755 AACCTTAAGGCGCCAAAGAACTAATGACCCAAACACAGACAACTATGCTGAGCT 2814
Db 760 -----ProSerProGlnAlaAspProLysAlaThrGluGlnVal----- 772

QY 2815 ATRACGTACAGTTTTCAGATATAGTGATATATATTCGAGAGAGATCTCTGGAAAGA 2874
DB 772 ----- 772
QY 2875 AACGGTACATGGTAAAGCTGCAAGACATTTGGAAAGTGTTTTGGTAATACATTAAG 2934
DB 772 ----- 772
QY 2935 TCACCTCAAGGCAAGAAATGATTAATATATGATGATGCCCCCAAAATTTTAAATGG 2994
DB 772 ----- 772
QY 2995 AGGAAATTTGGGAGACCTAATAGCCCAAGTATGGAGCCATGAAATGTGATATA 3054
DB 773 ----- 777
QY 3055 AATATTTGAGAGTAAATCGGACACCATACACAAAGTAGTTATTGGGATATAGT 3114
DB 778 ----- 778
QY 3115 GATCATACACATTTGATGATATATCCACAAATTAAGATGATGACCGAATGGCA 3174
DB 779 AsparGlu-----AsparPheGluGlu----- 787
QY 3175 GAATGCTACTGCAAGTGCAGAAAAGAGTATGATAGTTGACAGAGATTAAGAG 3234
DB 788 -----GluGluGluGluGluAspGluGluGluGluGluGluGluGluGluGlu 801
QY 3235 TGTAGCATAGATTAAGTGTCAAGCTGTACGAA-----GAGAGTGTACAGGT 3285
DB 802 ValGluGluGluGluThrAspGluSerAlaThrGluAlaValAlaProSerProGlu 821
QY 3286 TGTACAGATGCACAGACCTGTGAATGATATATATAGATTAAGATTAAGAAAGA 3345
DB 822 ThrThrGlu----- 824
QY 3346 CAATGAAATATATATACATTAATACAAAGATTTACATGACACAAATGTCTGT 3405
DB 825 -----AspGluValLysProAlaSerGluGluAspValLysVal 838
QY 3406 AGTAATAGTGTATGAGACCTTCAGTACTGCCAAAATCATATAGACAGATGTAT 3465
DB 839 CysSerIle-----ValAspLysAlaLeuLys 847
QY 3466 GAATTTTGTGCAATTTATACCAACAAATGTGSCAAAAGTAAAGTGTGTACTGT 3525
DB 848 GlyLysLeuAspAspAlaCysThrLeuLysTyrGlyLysThr-----AlaProThrSer 865
QY 3526 GATGAAAGTGTCTGCTATTTGGTACTAACCCAGCTATGAAAATGTGGAGCATATCTCCAT 3585
DB 866 TrpLysCysIleProSerGlyAsnAsnThrThrGluSerThr----- 880
QY 3586 GATACAGAAATTTTGATGATTTGTACGTACAAATGAGTTTGTGATGAAAAAGTAT 3645
DB 880 ----- 880
QY 3646 GGTAGGATTAACGAAATATATGCTTTAGAGATTAACACAGACCATGATGGCTGT 3705
DB 881 -----ThrLysPro-----GlyAlaAla 886
QY 3706 GGTGTTAAAGTGTGATGCAACCGACAGAGGTACAGATTAACGAAAAAGCGGAA 3765
DB 887 GlyThrProSerGly----- 891
QY 3766 GAAAGGATACGAGATTAACAGTGAATGATATACTTAAGAAAAAGATGGAAGAA 3825
DB 892 -----LysAspThr----- 894
QY 3826 CAAGTAGAAGATTTGCATCCAAAAAAGATAGTATGATATCCCATTTGCCAATGCGGA 3885
DB 895 -----Gly 895

QY 3886 AATATTAATTTAGTGGAGAACCTCGTGTCTATATCCCTTGAAGACAAAGTTATGC 3945
DB 896 SerIle-----CysValProProArgArgLysLeuLys 907
QY 3946 GTACATTTCTTC----- 3957
DB 908 ValGlyLysLeuHisAspThrAlaGlyGlyGluThrThrGluAlaLysSerGluGluThr 927
QY 3958 -----GCAATGATTAAT----- 3969
DB 928 SerGlyGlyGluLysThrProSerGlyAsnGluSerProSerGlyLysLeuProGlu 947
QY 3970 -----GAAATTAATAATTAACAATCAATCAAGTTATTAATAAGAGCTTCAATC 4017
DB 948 GlyProThrProGluThrThrLysGluThrProGluSerSerLeuHisAlaPheVal 967
QY 4018 AATGTGCAGCAGCAGAACATCTCTGATGATTAATTAAGTAAG----- 4068
DB 968 SerProProArgLeuArgArgPheLeuProThrHisLysPheLysGluGluThrLysAla 987
QY 4069 -----GATGGTGAAGCA 4080
DB 988 GlnHisGlyAlaGlyAlaThrGlyGluGluGluThrIleIleGlyThrLeuAspGlyGly 1007
QY 4081 AATGAACCTC-----GATTAAGATTAATAAGAGCAAAATCTCTCCGATTTTGAATCC 4137
DB 1008 GlnGluThrProAspLysLeuLysThrGlyHisIleProProAspPheLeuArgGln 1027
QY 4138 ATGCTTACACATTTGAGATTAAGATTTATTTATTTGA-----ACAGTATA----- 4188
DB 1028 MetPheThrThrLeuLysAspTyrAlaGlyAspIleLeuValGlyAsnThrPheIleVal 1047
QY 4189 -----TCAAAAGTCAATGCTGAGGAGTAACCTAAAGAGCAATAGATTTCT----- 4236
DB 1048 HisThrSerGlyLysAsnGlyAspMetGlnIleMetGluAlaIleGluLysIleGlu 1067
QY 4237 -----CTTTCAAAAATGTGACCAAAAATCTCCATATGAAAAACA----- 4278
DB 1068 GlnIleLeuProThrSerGlySerSerProSerProArgValThrGlnThrGlnHis 1087
QY 4279 -----CGCCAAAGATGCTGAGCAACATCATGATGATGAGTGAAGCT 4323
DB 1088 SerValGluAsnProArgLysThrThrPheAsnGluAsnLysLysIleThrProGluGly 1107
QY 4324 ATGCTATGTGCTACTGTA----- 4341
DB 1108 MetValCysAlaLeuThrTyrAsnThrAspThrProSerGlyThrAlaProThrGlnIle 1127
QY 4342 -----AAATTTGGGCAAAA-----AAGATGATTTTACGGAANAC-----TAC 4380
DB 1128 GlnGluValAlaThrThrLysLeuAspAspGluAsnSerLysAsnProLysIleProGluLys 1147
QY 4381 GGTTAACAACACCTCAAAATTTACTGAC----- 4407
DB 1148 LysTyrAspGlnValLysLeuAspAspThrSerAspAlaLysThrThrGlySerProVal 1167
QY 4408 -----AAAACACCACTTTGAGAGAAATTTGCCAAACACCCAGTTTAAACGA 4455
DB 1168 ProSerGlyGluLysIleThrProLeuThrAspPheIleSerArgProProGlyPheArg 1187
QY 4456 TGGCTAACGCAATGATGACGACGACTATGCTATACACGACGAAAAATTTTGAAGGATGTG 4515
DB 1188 TyrLeuGluGluThrProGlyLysThrPheCysLysGluArgLysArgLeuGluLysIle 1207
QY 4516 CAGGAAAAATGTAAAGTCA----- 4533
DB 1208 LysGluGluCysArgGlyAspArgThrGlnHisGluHisCysSerGlyLysArgLys 1227
QY 4534 -----AATGACCA-----TTGAAGTGTGATACAGAA 4560
DB 1228 CysThrArgThrAspAlaAspArgAsnAspLysPheValAspLeuAsnLysCys-----ArgAsp 1246
QY 4561 TGTAAATAGAAATGCGAGAGCTACGTTAAATATATATGAAA-----AAAAAAAAGAGTGATT 4617


```
Db 1247 CysHisIleGlnCysArgLysTyrArgLysTrpIleAspIleLysPheAspGluTyrHis 1266
QY 4618 CCACAAATATAATATACAGAGATGACGACGACAAATATTCGATACACACATC 4677
Db 1267 LysIleGlnLysLysTyrGlnGlyLutYrAspLysLeuThrLysAspLys 1283
QY 4678 GGTGTAATGGTTACAGACTACTGAGAACGAAACGACATTTACTGACAGAAATTT 4737
Db 1283 1283
QY 4738 ACTGCTAGTTGTGGTGAATAGCCGGAAGTCCCTGTGTGTACAGAAATATACATTC 4797
Db 1284 SerSerIleGlyLysP 1288
QY 4798 TTGAAAAACAGCGCTTACTATGATGCCACAAACATTTGTGTGACAAATTTATGAA 4857
Db 1289 AsnAsnCysCysLysAspIleGlnLysHisLysSerAlaAlaValPheLeuLys 1306
QY 4858 AATGACGACAAATATATCTAATCAATTTGAGTAAGTAAGTGCAGAAAGATTAGTAAGAG 4917
Db 1307 Gln 1310
QY 4918 GCAACACAGCTGCTATTAAGTGCAGAAACAAAGTCTTAATACATCAATACCTGAAA 4977
Db 1311 CysLysAsnGlyIleThrSer--GluAsnLysGly--AsnGlnGlnAspGlnLeuAsn 1328
QY 4978 GAATTAATCTGAGATGCTGCTTT--TCCTTCCTGCTGCTGCTGCTGCTGCTGCTGCT 5033
Db 1329 LysLeuAspPheAspLysLysIleProGlnThrPheSerProSerThrTyrCysLysLacys 1348
QY 5034 AATGATGCGCAATTAATACATCAATCAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 5093
Db 1349 1350
QY 5094 GATGGAAGTGGCGGACGACGAGAGGTACATTTGGTCAATACTACAAAGAAAAAGAA 5153
Db 1351 TyrGlyValAsnCysAsnGly 1360
QY 5154 AAAAGGAAATATAAAGCGTGGATGCGACAAATATCTTATGAGTCCGCTTGTAG 5213
Db 1361 GlyArgGlyGlyThrAsn--GlyCysThrThr 1370
QY 5214 TGCATGAATAATAGTTTATTAAGATTAATTAATTAATTAATTAATTAATTAATTAAT 5273
Db 1371 1378
QY 5274 AGATGAAAAACAAAGACCGAGAGAAATTTGACAGAAATTTTAAACAAAAATGGAACATC 5333
Db 1378 uasnasp 1380
QY 5334 AGTTGGCAAGGAGTACTACTACAGAG--AATCCCGTACTACTGCGCG 5384
Db 1381 LysGlyAlaAlaSerThrIleSerIleLeuIleAsnAspGlySerThr 1396
QY 5385 AAAATTTTTCGACACGAAATTAAGCAATGTGTGTGACCAACCAATGATCCGGGTACAA 5444
Db 1396 1396
QY 5445 ACGTGTAGGAGATGATGATAGTGAATATGTCAGCAAGAGATGATGATGATGATGATGAT 5504
Db 1397 AsnGlyAlaAlaThrAsnGlyThrThrGlyThr--ThrAspGlnThrLeuLeuGln 1413
QY 5505 ATGTGGTCTGTACTCTCAGATGATGATTAATCTATG--GGGAAAAATCCGCGATGA 5558
Db 1413 uCys--SerAspLysTyrAlaPhePheLysGlyLeuArgGlyGlnGln 1428
QY 5559 AGTACTGCGTATGATTTCTTCGATGCTTTGCCGAATGGGAGCAATTTTTCGAAACA 5618
Db 1428 uTPTrhCysGlnLys--LysTyrGlyValAsnGlnCys-- 1440
QY 5619 TAAAGAAAGCAATTTGAGAAATTTGAGGCGGTGTAATGATTAATTAATTAATTAATTAAT 5678
Db 1441 1453
QY 5679 TGAAGATTAAGAAAGAAATGTCACAGATCGCTGTACACAAATTAATTAATTAATTAAT 5738
Db 1453 sasP 1463
QY 5739 GTGGAAACACAGTATGAAAAAATCAAAAAATATGATGATGATGATGATGATGATGAT 5798
Db 1463 gtrPheLeuArgTyrPhe--ValHisAspTyrAsnIleLeuLysHisLysIle-- 1479
QY 5799 TTCGAGATCCTGTGCGCAAAAGATGACAGAGACCTGCGAAATTTTAAGCAACAAT 5858
Db 1480 1483
QY 5859 AAAAAAATTTGTGAAAAATTAAGTGAATGATGATGATGATGATGATGATGATGATGATG 5918
Db 1483 eLysLysGlnLysGlnAspLysThr--GlnHisLysCysIle-- 1496
QY 5919 CACACAGGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5978
Db 1497 1501
QY 5979 AAAAGAGTGAAGGAAAGTGAATTTGCAAGTGCACAGAGTCCACAGCGTGAAGAG 6038
Db 1502 1507
QY 6039 GGAACACCGCTCACACAGGATATCACTGATATCAAAAACGACGCGATCGAAAAAGAAC 6098
Db 1507 1507
QY 6099 GAAAAACCGCGCTTACAAAAACAGCGAAAAAAGTGAAAAATCTAACACAGAAATGCG 6158
Db 1507 1507
QY 6159 AGCACAACACAGAACCCGACAGACGACACAAACAAACAGAAACATCAACAGACAC 6218
Db 1507 1507
QY 6219 AACACAGAAATCTGACGTGGGACAAATGTAAGGCGATCTTTCGAAATTAACACATAG 6278
Db 1507 1507
QY 6279 CAGGCGTGAATAGAGGTTGTAATCCAAAAACGTATGCAATATTCCTTAATGCGGTTG 6338
Db 1508 1511
QY 6339 TATGTAGTAAGTCTTAAGAAATGAAATGSCATATGATGCTCTTAAGGAAAAA 6398
Db 1511 uIleLysGly--AsnGlnTyrGly 1518
QY 6399 ATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6458
Db 1519 1530
QY 6459 TATAAAGAGCGCTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6518
Db 1530 pasPlysGlnThr 1534
QY 6519 TATTAATGAAAAATCTGACAGCAAAAAATGAATTCGAAATTAAGAACATTCAGATGAAT 6578
Db 1534 1534
QY 6579 TAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6635
Db 1535 1552
QY 6636 TATTTCTATGATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6695
Db 1552 ptyrLysLysAlaGlnLysValVal 1560
QY 6696 AATTAATTAAGAAAAACAGATTAATAAATAAAGATGAAGAAATTAATTAATTAATTTGGGA 6755
Db 1561 1570
QY 6755 1570
```

QY 6756 GAAAAATAAAATTTATTTGGAGGAATGATATATGATTAATCAATCTCACAGA 6815
DB 1571 -----GlyCysThrGlyHis-----As 1576
QY 6816 CGAA---AACGAAAAAGAAAAATTTAGATATATTTACCAATGACATGACCAACT 6872
DB 1576 pGluCysSerGluLysGluLysGluLys-----AspHeiLeuThrSniLe 1593
QY 6873 GACGCTTCCTGAGAGAGTTGTAATAAAGCCCAATTTTGGAGTGTTCACAGAAATG 6932
DB 1593 uLieserGluLeuGlnAspLysIleThrSerCysGln----- 1605
QY 6933 GCGAGAAAGATTTTGTAAATAGAGAGAACAGTTGTTAAATGAGCGCGCTGTAA 6992
DB 1606 -----AsnLysHis----- 1608
QY 6993 GGAATATGAGTGTATGTATGATATGACGCTAAGACACAGAAATGCGAGCGCTGTCT 7052
DB 1609 -----AsnProAsnGluLysThr-----AlaCys-- 1616
QY 7053 AACATATCAAAATTTTATTATAGAGAGTGAAGTGAATGAAAGACAAAGAAAGTT 7112
DB 1616 ----- 1616
QY 7113 CAAAAAGATTAAGATGCAAAAAAGTATAGGATTATCTTACTGAAAAGACATAGA 7172
DB 1616 ----- 1616
QY 7173 GAAGCAACATGTCTCATGAAATATTTAAACATGAAATTTAAAGATTTGTGGCAATTA 7232
DB 1616 ----- 1616
QY 7233 GGATGTCTGTGTATGCAAAAACTTCTTCAACAATACCAAAAAACACAAATCACA 7292
DB 1617 -----AspProHeProSerProThrProGluLuh 1627
QY 7293 ATCATCCGATGCTATGATATGCCAGAAATCGTGATATGTTCTGAAAGATTTAACA 7352
DB 1627 rAspProLeuAspAspAspThrProAspProLeuAspAspAspGlnHisThrGluGlnPr 1647
QY 7353 GGTGAGTGTCTGAACTTCAAAAAAGGATCTATGATTCATCAAAAAAATTTACTGA 7412
DB 1647 oLysPheCysPro-----Pr 1652
QY 7413 ACCTAAATATACCTGATGATGTTGTAGAGAAAGCAGCATATTTATTTATTAAGACAGCA 7472
DB 1652 oProProProProMetIleThrCysValGluLysIleAlaLysGluLeuArgValGluAlaG 1672
QY 7473 AATTAATATGATATTTACTCTGAAAGAAAA----- 7503
DB 1672 uGluLysIleAsnAsnGluLeuLysGlnLysAspPheAsnGluLysCysAsnAs 1692
QY 7504 -----TTTATACCTATGAGTGTACAAAG-----GAAAAGAGAAAG 7538
DB 1692 nValLysLysLysAsnGluAlaValIleGluLysGlnLysCysLysPheGluLuhThrTy 1712
QY 7539 TAAAAATAGTGTGATATATATATATCTTCGATCCTTAAGAAACCTTATGACACCTGATA 7598
DB 1712 rGluAsnSerValAsnAsnIleAsn----- 1720
QY 7599 AATATATAGAGAGAAAGCCCTTGTAAATAATAGAGAAATCGTTTATAGGATGATTA 7658
DB 1721 -----AsnLysCysLysAspAsnGlnAsnGluArgPheLysIleGluG 1735
QY 7659 TGAATGGAATGTTACAAAAATTCAAAGTCTTATCAAGAG-----AAAAAGAGATATG 7715
DB 1735 nLysStrp-----AsnHeLysStrpIleGluLysIleLeuLysAspLeuLysI 1751
QY 7716 ACCTCCAGAGAGAAACATATGTCTTAAGAAATTTAGATGAATTTAAATTTAAAGACT 7775
DB 1751 eProProAlaGluGluHisMetCysLeuAspAspLeuSerMetLeuGluArgThrThrI 1771

QY 7776 TAAAGTATATATTTATCTCTTAAAAATGGTGTGCACTGCCAGAAATGAGGAATAGA 7835
DB 1771 eSerAspSerSerLeuLeuLysIleGlnGluAlaLysSerGluArgAsp 1791
QY 7836 CATTAATAAAAAATCTTCACTCAGAGACGGCGCCAGATGAATCCAAATTCGATCTACT 7895
DB 1791 pIleIleArgLysLeuLeuGluGlnAsnSerCysAspGluHisArgIleCysAspAlaLe 1811
QY 7896 GAAATATAGTTCGCTGATCTGGGATGACATAGTTAGAGCAACAGATATGTTGCAATTCG 7955
DB 1811 tLysTySerPheAlaAspLeuLysPleIleArgGluArgAspLeuThrAsnLysAs 1831
QY 7956 TGGTACTTACTCCCGTAGAAATAAATTAATTAAGTTTGAATACATATATGGA 8015
DB 1831 nSerLysGlnLysGluLeuGlnLysArgLeuGlnLysArgIleAsnIleTyAsnL 1851
QY 8016 ATCGAATAAT-----AAAT-----AAGGTACAAATTAATACAGATGTACAAAC 8063
DB 1851 sLeuGlnAsnAspLysAsnLysTyGluLysAspArgProLysTyLeuGlnLeu----- 1869
QY 8064 GTTTCGCTGCTGCTGGGATGCTAATAGAAAGATATTTGGAAGCAATGACGTGCA 8123
DB 1870 ---ArgSerAspTrpTrpAspAlaAsnArgLysHisIleTrpAsnAlaMetThrCysAs 1888
QY 8124 AGCACCAGAGATGCAAACTTTTACAAAAGAGAGATGAT-----GG 8168
DB 1888 nAlaProAspAspAlaLysPheLeuLysAsnProAsnAspThrSerGlySerSerSe 1908
QY 8169 ATTGAACGCATTAATTAATACAGATAGTGTGACATTAAGAGAGATCCACCTGTGA 8228
DB 1908 rSerLysGluLysIleMetThrThrHisSerAsnGlyTyAspLysGluProProAspTy 1928
QY 8229 TGAATATATACCTCAACGGTTCGATGATGATGATGATGATGATGATGATGATGATG 8288
DB 1928 rAspTyTrpIleProGlnProPheArgTrpMetGlnGluTrpSerGluSerPheCysLysLe 1948
QY 8289 ACTGATGAGAGAAATTTAAATAATATCATGTATGACTGT---AAACATCTGA 8345
DB 1948 uLeuAsnGluLysMetGlnIlePheGluLysThrCysGluLysGluLysAsnSerI 1968
QY 8346 CAGATGCAAGATGATTAATGATTAATAAGTGTGAAACAGTGTAAACAGATGTCAAG 8405
DB 1968 eThrCysGluAspAspArgAsnGlyThrAsnGlyAsnGlyLysAsnGlnLysGluL 1988
QY 8406 ATTAATAAATTTTCTTAATAATGCAAAATCTATGATGATGATGATGATGATGATG 8465
DB 1988 sTyLysLysLysIleHisAsnTrpLysLeuGluPheAsp-----LysTyTy 2004
QY 8466 AGAATGTATGAACAACAATATAT-----ACAAAATCTTACTTATGATCA 8513
DB 2004 sGluIleTyAsnGlu---IleTyAsnAsnLysAspSerLysIleAsnSerAsnGluTy 2023
QY 8514 TGTTCAAAATTTTGTCAAAAGTGTGAAACTTTAAAGTGAATG---TCGTGTGAG 8570
DB 2023 rPheLysLysPheLeuGluLysLeu-----LysAspLysCysLysGluLeuLysAsnSe 2040
QY 8571 CTTTTCGATATATCTCATGAAACAAGTAAGTGTGATTAATTAATTAATTAATTAATGA 8630
DB 2040 rSerAspLysCysIleAspGlnAlaThrHisCysThrLysTyLysPheSer---AsnSe 2059
QY 8631 TGGTCTTCATATATACGAACATATGCTTTCGAAGAAACCAAAAGTTAATAAGAAC 8690
DB 2059 rGluAsnLysAsnHisAsnAsnTyGluArgPheLysAsnProProGluGluTyLysAl 2079
QY 8691 TTGCACTGTGACACTCTTCAAGATCCATGATGATTAATTCGTCACGATGAA----- 8745
DB 2079 aCysLysCysAspAlaPro-----AspProLeuAspAsnCysProLysAspSerAlaTh 2097
QY 8746 -AACAAAGATGATGTAAAGAAATTAACAACCTTTTACCTTCGCAAGATATGATTA 8804
DB 2097 rTyGluLysAlaCysAsnThrLeuLeuProThrLysLeuCysGluSerLysThrPheAs 2117
QY 8805 TAATAATCTTGATTAATGGAACGCATACCTGTCTTAATAGTTCAAGATGAACAAG 8864

```

Db 2117 nasnaaspaspserrtrpserphrSerPheValGlnThrSerProargspasnThrG1 2137
QY 8865 TGTATGATTCCTCAGAGAGACATTTATGACAGACCTATGCCTCATTAATTA 8924
Db 2137 yValLeuValProProKarghArgGlnIleCysLeuLysAsnIleThrThr---LysLe 2156
QY 8825 TAGAAAAGTCATAAA---GAAATTTTAAAAAAAACCTTACTCTGCTTCAAGTCA 8981
Db 2156 uArgSerIleGlnLysIleAspSerPheLysAlaGlnLeuMetThrSerAlaLysAsnG1 2176
QY 8982 AGGCAATGTGTGGTCAAAAATTAATTAATGGAAGAAGATTGCTTGAAGCAATGA 9041
Db 2176 uGlyLysLeuLeuCysGlnLeuLysLysAspArgAspValThrLeuGlnAlaMetLe 2196
QY 9042 ATATAGTTATCGAATTATTCGATTAATTAATTAAGAACATGATGAGCACTTCATT 9101
Db 2196 sTyrSerPheLysArgLysArgLysArgLysArgLysArgLysArgLysArgLys 2216
QY 9102 ATCTGAAAAAATTAATAAATTTT-----GAAACATCAAAATGAAGC 9143
Db 2216 oLeuAspLysLeuLysThrLysLeuAsnValLeuLeuLysGlyAspGlyThrAsnGlnI 2236
QY 9144 AACCGAAATCGTAAACATGCTGGGAAAAATATAGACGTCAGATATGGCAGCTATGT 9203
Db 2236 eLysGlnAspArgGlyLysTrpTrpThrGlnAsnArgThrArgValThrPheIleAlaMetLe 2256
QY 9204 ATGTGATATTAATTAATGCTACTCAAAAGTAACTAGATGAGAGAGAGCTCAATAC 9263
Db 2256 uCysGlyLysLysAlaAlaGlyLysIle-----GlnGlnArgAspCysSerLeuPr 2274
QY 9264 AAAGATGAGAAACATTAATCAGTTCTCTGTTGTTAATGAATGGGCAACAAAGCATG 9323
Db 2274 o---AspAspAsnThrHisGlnPheLeuArgTrpPheArgLysIleLysPhe 2293
QY 9324 TAAGCAAAAGAAACATGTAAGTATGATTAATAAACAATGCTCTGTTCAAGCAAGA 9383
Db 2293 sAlaLysArgGlnLysLeuPheAsnGlnValLysArgLysCys-----AlaSerAlaG1 2311
QY 9384 TAATTTGAAGGCTGAGATTTATTAAGACAACTGATGTCAG-----AATGATAT 9434
Db 2311 nCysIleIleGlnLysGlyLysThrIleAspProProValCysGlnGlnAlaCysThrGlnTy 2331
QY 9435 TGAAGAAATTAATTAATGCTGAATTAATTAATAAATGAAATGAAATCTAAATATA 9494
Db 2331 rArgAspLysIleThrArgLysIleGln-----G1 2341
QY 9495 ATATAGCAATTAATAAAGATCACTCTCAGTAATATAGACAATAAACCATTGA---GA 9551
Db 2341 uTyrArgLeuLeuAsnLysGlnLysArgLysArgLysArgLysArgLysArgLysArgLys 9552
QY 9552 AAATGTTGAGTCATATTAATAATCAAAAGATTCATGACGCTTGGAGTAAATGATAT 9611
Db 2361 rLysAlaProGlnLysPheAsnAspLys-----CysAsnAspLysCysAsnCysLe 2378
QY 9612 AAATGAATTAATCTAGACAAACAAATAATGAATAATGAATCAAGTCAAGTCAAA 9671
Db 2378 uSerLysLysIle-----AspIleGlnLysLysIle 2389
QY 9672 AAAATTTATATCTGTTATATTTTGTGAAGATGAACACACAAAATCATGACTAGA 9731
Db 2389 sAsnMetLysTrpSerPhe-----AspAspAsnAspLeuLysAsnLysCysIleCys 2406
QY 9732 TGAAGAAATTAATAAAGAGAGACAGTTCGCTAAAGCACTATATTTCTTATACAC 9791
Db 2406 sArgGlnIleLysProLysArgPro-----ProLysLysAlaLysProGlnGlnG1 2423
QY 9792 CCATGTAGATCTTCTTCATACAGACCTTATTTCTCAACACATGAGTACACAAATATGA 9851
Db 2423 uHisThrProSerGlnGlnLysAspThrProProProLeuProProLysProAspAspLeuPr 2443
QY 9852 TCCT-----AAAATGATATATGAAGAAGTATCTCTGTTGAT 9893
Db 2443 oProProAlaGlnGlnLysProPheAsnArgAspIleLeuGlnLysThrIleProPheGlyI1 2463
QY 9894 TGTATGCGCGGTAGCTTGTAGACCGCTTCATTTGATGAAGAAATAAATCAATGCTGCT 9953
Db 2463 eAlaLeuAlaLeuGlnLysSerIleAlaPheLeuPheLeuLysLysLysThrLysSerSerVa 2483
QY 9954 G---GACTTGTTCGCTATACATAATACCCGAGAGAGATGAGATGCCCTAGCTTGA 10010
Db 2483 lLysLysLeuPheGlnIleLeuIleIlePheLysSerAspLysArgIleProThrLysLe 2503
QY 10011 ATCCAAAAATAGTCATATACCATATATAGAGTGGTCCATATATTAAGCAAAACATATATA 10070
Db 2503 uSerProAsnArgLysIleProLysThrSerGlyLysArgLysArgLysArgLysArgLys 2523
QY 10071 TATGAGAGAGATCT-----AGTGAAGATGAGATTAATATATGCTGGACTATAC 10121
Db 2523 rLeuGlnGlnLysSerGlyThrAspSerGlyLysThrAspHisLysr----- 2538
QY 10122 TTCCTGATATTTACT---TCATCCGAAAGTGAAGTATGAGATTAATGATAT 10178
Db 2539 ---SerAspIleThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 2557
QY 10179 ATATGTACAGTAGTCTCTAATAATTAACATGTATGATGATAGTACTAGACATCAAA 10238
Db 2557 eTyrValProGlnLysSerProLysLysLysLysLysLysLysLysLysLysLysLysLys 2577
QY 10239 AAGGATATACCAAGTCAAT-----GATACCAAGT-----AA 10271
Db 2577 yAsnAsnThrThrAlaSerGlyLysAsnThrProSerAspThrGlnAsnAspIleGlnAs 2597
QY 10272 TGATACACACAGTCAATGATTAATGATGATGATGATGATGATGATGATGATGATGAT 10331
Db 2597 nAspGlyLysProSerSerSerLysIleThrAspAsnGlnLysLysLysLysLysLysLys 2617
QY 10332 TGTATCTCAATATTA-----CCAAATACAGAAACAAATTAATTAATTAATTAAT 10382
Db 2617 eIleSerAsnMetLeuGlnAsnGlnLysProAsnThrGlnPro-----AsnMetLeuGly 2635
QY 10383 AGATATTCATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10442
Db 2635 rAsnValAspAsnAsnThrHisProThrThr---SerArgHisAsnValGlnLysLysPr 2654
QY 10443 TTTTATATATCTATTCATGATAGGATTAATTAATGATGATGATGATGATGATGATGAT 10502
Db 2654 oPheIleMetSerIleHisAspArgAspLeuLysGlnGlnLysLysLysLysLysLysLys 2674
QY 10503 TAATATGAGTAAATTAATTAATGATTAATTCATGATGATGATGATGATGATGATGAT 10562
Db 2674 lAsnMet-----ValAsnAsnAspIleProIleSerAlaArgAsnGlnLysAsnTySe 2691
QY 10563 AGTATATGATTAATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 10622
Db 2691 rGlyIleAspLeuIleAsnAspSerLeu---AsnSerAsnLysValAspIleLysArgGln 2710
QY 10623 GATATG 10628
Db 2711 LeuLeu 2712

```

```

RESULT 10
Q904A2
ID Q904A2 PRELIMINARY; PRT: 1685 AA.
AC Q904A2:
DT 01-MAY-2000 (TRIBLEL. 13, Created)
DT 01-MAY-2000 (TRIBLEL. 13, Last sequence update)
DT 01-MAY-2002 (TRIBLEL. 20, Last annotation update)
DE Variant surface protein pFEMP1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IT;

```


|||||
Db 594 Ileasnncysasnlygascyslyscyltlystrlylgllylscylgllyl
QY 1924 GAATGGAATAGTAATGAATCTGTTCACAAATAAAGAAATATACGAATGATATAT 1983
Db 614 Glutprlyllyllylvalylstgyllylscyllylscyllylscyllyllyl 633
QY 1984 AGTAATATTAATATCTTTTGAAGCTTATTTTGAAGTATGAATGAATCTGACAA 2043
Db 634 Valileuyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 653
QY 2044 GATGACCAATATGGAAGAATGTAATGAATGAATGAATGAATGAATGAATGATTC 2103
Db 654 Glutlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 673
QY 2104 AATTTGGAATATTAATGAATGTAATGAATGTAATGAATGTAATGTAATGTAAT 2163
Db 674 Aspraplyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 690
QY 2164 AAAGAAATCTGCAATGTAATGAATGAATGAATGAATGAATGAATGAATGAAT 2211
Db 691 Leuyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 710
QY 2212 -----GAAACATCCCATATATGCAACAAACCCGCTG 2244
Db 711 Prohlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 726
QY 2245 GTTAAACCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2304
Db 727 Glylyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 743
QY 2305 AAAGAGATGATACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2337
Db 744 Gluargaplyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 763
QY 2338 GGTCTTCAT---AAATGAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2394
Db 764 Gluargaplyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 783
QY 2395 ACAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2451
Db 784 Glyseraplyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 803
QY 2452 GTTGTCTTCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2505
Db 804 Argglyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 820
QY 2506 ACAAGATTTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2565
Db 821 Aspraplyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 839
QY 2566 GAAGATGTAATATGCTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2625
Db 840 Seraplyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 859
QY 2626 CAAGAGGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2685
Db 860 Gluargaplyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 879
QY 2686 TTGGGAGATGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2745
Db 880 Leuyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 899
QY 2746 GAAAGAAATTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2805
Db 900 Asnlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 916
QY 2806 TGTCAGATTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2865
Db 917 Cysarglyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 936
QY 2866 TGCGAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2925
|||||
|||||

Db 937 Trpasplyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 956
QY 2926 ATTCATTAATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2985
Db 957 Ilelyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 976
QY 2986 TTAATTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3045
Db 977 Thrlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 996
QY 3046 TGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3105
Db 997 Cysalialle-----Glnaplyllyllyllyllyllyllyllyllyllyl 1007
QY 3106 GGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3165
Db 1008 -----AspGlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1024
QY 3166 GAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3225
Db 1025 Gluargaplyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1044
QY 3226 TGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3285
Db 1045 CysSerGlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1061
QY 3286 TGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3345
Db 1062 Cysasnlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1081
QY 3346 CAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3405
Db 1082 Gluargaplyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1101
QY 3406 AGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3465
Db 1102 Alaasnlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1121
QY 3466 GAATTTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3522
Db 1122 Asnheleuplyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1141
QY 3523 AGTATGAAAGTGTCTCAT---GCTACTAACACACAGATGAA 3564
Db 1142 Hisprolyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1161
QY 3565 AATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3624
Db 1162 Thrlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1181
QY 3625 TTTTGTGATGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3681
Db 1182 Pheyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1201
QY 3682 CCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3741
Db 1202 Protyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1220
QY 3742 ATTAATAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3801
Db 1221 -----Leuyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1231
QY 3802 CTTAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3861
Db 1232 IleSerGlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1249
QY 3862 GGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3921
Db 1250 Ala-----Lyslyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1267
QY 3922 CCCCCTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3981
Db 1268 Protyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1287

```

OY 3982 TTACATCAAGTATATTTAAAGAGCTTTCATCAAAATGCGACGACAGAAACATTC 4041
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1288 -----GlucInaspLeuArgGluAlaPheIleYSerAlaIaIaGluThrPhe 1304
OY 4042 TTCTCATGTATATATTAAGTAGAGTGAGGAAAGCAACTCATTAAGAAATTA 4101
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1305 LeuLeuArgGluThrYTrpYTrpSerLysAsnValGluAsp-----AspLysIleLeu 1321
OY 4102 AAGAAGCAAAATTCCTCCGCAATTTTGACATCGATGTCACATTTGAGATTA 4161
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1322 HisArgAspMetLeuProGluPhePheArgSerMetPheYThrPheGlyAspTyr 1341
OY 4162 AGAGATTTTATTGGAACAGATATATCAAA-----GGTCATGTGAGGAGATAA 4215
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1342 ArgAspIleCysLeuAspThrAspIleSerGluLysIleAlaAspHisAspValThrThr 1361
OY 4216 CTTAAAGACCAATAGATCTCTTTCAAAAATGTGACCAAAAATCTCTAATGAAAA 4275
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1362 AlaLysLysLysIleThrAlaValPheGlnLysIleGlySerLysThrThrAsnGlyLys 1381
OY 4276 -----ACAGCGCAAGATGTGACAGACAGATAGCATGATATGGAAGCTATG 4326
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1382 LysValLeuGluArgGluGluGlyTrpLysGluTyrGlyLeuSerIleTrpLysGlyMet 1401
OY 4327 CTATGTGACCTA---GTAAAAATTGGGCAAAAAAAGATGAT-----TTT 4368
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1402 LeuCysAlaLeuSerYTrpAsnThrGluThrLysLysMetAspLysValArgThrTyr 1421
OY 4369 ACCGAAATCTACGGTTACAAACAAACGTCAAATTTAGTACAAAAAGCACCTTTGGAGAA 4428
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1422 LeuMetLysTrpIleTyrLysAsn-----AsnAspIleLysGluTyrLeuGluGln 1438
OY 4429 TTTCGCAAAAGCAACCCGATTTTACGATGGCTAACCAAGTATGACGACGATCTAT 4488
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1439 PheAlaSerArgProProPheLeuArgTyrValThrGluTrpLysAspPheValLys 1458
OY 4489 ACACGCAAAAATATTGTAAGATGTGACAGAAAAATGTAACTCA----- 4533
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1459 AsnArgLysLysGluLeuValSerLeuLysLysLysCysAspSerCysThrLeuArgAsn 1478
OY 4534 -----AATGACCAATTTGAAGTGTACAGAA-----TGTAATTAAGAA 4572
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1479 AsnGlyThrSerAsnLysThrCysAspAspAsnGluAsnCysGlyAlaCysLysThrGln 1498
OY 4573 TTGAGAGACTACGTTAAATATATGAAGAAA-----AAAAAGAGTGGATCCACAAGATTA 4629
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1499 CysGluLysTyrLysTrpMetGluArgTrpLysHisLysSerGlnLysLys 1518
OY 4630 TATTACAGAGATGACCGCAAAAAGATTTGATAGACACACATTTGGTATGTT 4689
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1519 LysPheGluLeuTyrLysAsnSerAlaThrTyrAsnAsn-----GlyLeuAlaVal 1535
OY 4690 ACGCATATATACGACAGCAAGATTTACTTAACAGAAATTTACTGTAATTTG 4749
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1536 LysGlu-----AlaAsnSerGluThrTyrLysAsnAspProGluValThrGluAla 1552
OY 4750 GGTGATAGCTGGAAGTGCCTCTGTGTACAA-----AGAATATACAAATTTGTTA 4800
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1553 AsnSerAlaLysHisAlaArgAspTyrLeuLysThrGlnLeuLysMetIleCysThr 1572
OY 4801 GAAAAACAGGCTTACTATGATGCCGACAAACATTTGGGTGACAAAAATTTATGA 4860
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1573 AsnGlyAsnThrTyrLysAsnCysAspTyrThrCysMetAsnThrSerSerThrAsn 1592
OY 4861 GAGCAGC-----AAATATACAAATTTGAGTAAATAAATAGCAAGCA 4905
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1593 SerGluMetProAlaSerLeuAspTyrThrProSerGluTyrLysAspLysCysAsnCys 1612
OY 4906 TTAGTAAAGAGGCAAC 4923
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1613 ValProAspPoluLysSer 1618

```

```

RESULT 11
O8T5G0 PRELIMINARY; PRT; 2658 AA.
AC O8T5G0;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Erythrocyte membrane protein-1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JDP8;
RA Chactopadhyay R., Pillai C.R., Chitnis C.;
RT "Identification of a domain responsible for binding to intercellular
adhesion molecule-1 from a Plasmodium falciparum field isolate."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBD databases.
DR EMBL; AY028643; AAK49742.1; -.
FT NON_TER 2658
FT 2658
SQ SEQUENCE 2658 AA; 301326 MW; AEE33BC9167E7B7 CRC64;

Alignment Scores:
Pred. No.: 8.7e-137 Length: 2658
Score: 2869.00 Matches: 872
Percent Similarity: 38.52% Conservative: 373
Best Local Similarity: 26.98% Mismatches: 909
Query Match: 14.75% Indels: 1078
DB: Gaps: 122

US-10-087-013-1 (1-10628) x O8T5G0 (1-2658)
OY 91 AGTCAAAAGTGCAGAAATGTTTGAACGTTATGCCAAAATATAGACATCATCA 150
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 11 SerGlyGluThrAlaLysGluLeuLeuAspArgIleGlyGluGluVal----- 26
OY 151 AATATGCAAAAGACATGTGATTCGTTGAAGGGAT-----TTG 192
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 27 ---TyrGluLysAlaHisGluAspAlaLeuLysGlySerMetGlyLeuLysGlyThrLeu 45
OY 193 ACGAAAGCAGAAATTCGTCGTGTCCTTTCACCCAGTAATAAGCATATTTATAT 252
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 46 SerGluAlaLeuPheGluLysAlaProGluGlyLysGlnThrSerGluAsp----- 62
OY 253 CCATATCATGTAATTAGATCATTAAGAACATTAATTTACG-----TATGATGAT 306
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 63 -----ProCysAspLeuAsnHisGluTyrHisThrThrValThrSerGlyTyrAspLys 80
OY 307 GTGAATTTGAGACATCCTTGCATGTGAGAACAAACGATTTGATGAAGTAAGAA 366
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 81 GluAsn-----ProCysLysAspArgProGluValArgPheSerTyrThrGluGly 97
OY 367 TCTGATGTGGA---AATAAATACGTAATTTAAAAAGAAAAATATGCTATAGCTGT 423
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 98 AlaGluCysAspLysSerLysIleArgGlySerAsnSerAsnLysAspGly---AlaCys 116
OY 424 GCGCCACATGAGAGACGACATATGTGATTAACATTTGGAAGCTCA---AATGATATA 480
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 117 AlaProPheArgArgLeuHisLysCysAspGlnHisLysGlnHisLysHisAspLys 136
OY 481 AATACCAAAATTTGATGATTTATTTGGGAATGTAAGTACTAGTACAGCAAAATAGCAAG 540
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 137 IleThrArg-----HisAsnLeuAlaAspValCysGluAlaHisLysPheGluAla 154
OY 541 GATTCATTT-----GTAAATATATCATCCATAAAGAACT 576
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 155 SerLeuGluLysTyrArgGlyGlnTyrGlnLeuAsnAsn----- 168
OY 577 TCAGACGCT-----TGTAAGTCTCTTGCAGAAAGTTTGCAGATATAGCTGAT 624
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169 SerAspValAsnIleAsnIleCysThrGlnLeuAlaArgSerPheHisAspIleGlyAsp 188

```


OY 625 ATTGTAGAGAAATGATATGTTTAAACCAATGTCATGACAAA----- 669
Db 189 ILevalrgegllylsaspleupheleugllyasnthryglseralaglathrgly 208
OY 670 GTAGAAACGGGTCCGAGAGGTTTTCAGAAATATACATGAT-----GCA 714
Db 209 Leuglunsllyleuylsgluliephneglnlyllehisserspvalthrlysglygly 228
OY 715 ATGCAAGATCAAGTAAATAAATGATTAACATCCGATGATCTGCAATATTAATTA 774
Db 229 Thrasnelygluleuylslysarplyglulysaspolyln---Asnpherygluleu 247
OY 775 AGAAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 834
Db 248 Argglunslprtrprthrprthrprthrprthrprthrprthrprthrprthrprthr 267
OY 835 TCATATATATATGAT 894
Db 268 G1YThrserasplysryrphearpnanthrycysershasnplysAlaelythrsergly 287
OY 895 AATGCGGCATTAACAGAGAAAGTTCCATCAATTTAGATATGTCCTCAATATTTA 954
Db 288 Lyscysarplycysasnspasnlnvalprothrtyrphesaptryvalproglinlyleu 307
OY 955 CGTTGGTTCAGCAATGGGAGAGAGTTTCCGAAAAGAAATATTAATGAAAAG 1014
Db 308 Argtrphegluclutrrpalaaglunsphepcysarplyleuarglyshislyleuun 327
OY 1015 GTCACAGACTCCTGTCGT-----AATGACAAAGAACCTTATATGATGATATGCA 1068
Db 328 AlaAlasnllyscysarplysproasnlyvalglulys---Tycysasnpleunsnarg 346
OY 1069 CATGATGTGACGACACTTTTGAAATAAAGATATTTTCATTTGGATATATAGTCTACT 1128
Db 347 TyraspcysglulysThrAlaserglyLysHisasrphheglunspasvalyslys 366
OY 1129 GACTGTTCACATAAATGCAAGTTTGTGAAGTTGTTGGAATGCAATCAATGATTT 1188
Db 367 AspCysglinlysercysAlaProphevalasrtrpilleaspsnclunlyleuun 386
OY 1189 AAAAAAAGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAA 1248
Db 387 Leulysglinargasnlysglyglulysglumetlnlysty---Thrasnelythrlys 405
OY 1249 TTTGTCAATATATATATATATAGTGA----- 1272
Db 406 -----AsnserlysarplyslysglyprogllyAlaelyLysSerasntryasp 420
OY 1273 ---TATTAACAATTTTATGAAAAAATGAAAGCAATATGCAATATGACACT 1329
Db 421 G1YTYrG1ulysarphethyrasnllleu---glulyslytyrAsnAsnValasplys 439
OY 1330 TTTTAAATTTACTTAATGAAGAAAGATATGTAA-----GCA 1368
Db 440 PheleuaspleuLeuAsnAspOluThrThrcyslyslsAsnAsnlnlleglunlgl 459
OY 1369 GGATTAACAGAGAAAGATATCTTTACTTAACAGTGTGATGAC-----AA 1419
Db 460 G1YlnlileasnphelyAsnValasnserylylsasnserhaspaspaspserasn 479
OY 1420 GGGATATTTATCGTTTCGAATATGTCAGTGTCCGAGCTCGGGGCAATATGAT 1479
Db 480 LysThrphcysarplyrthrtyrcysglinalacyprotrpcysgllyalaglunlasp 499
OY 1480 GGTATTAATAATC---ACACACAATTCAGATATGATGTCGACGTGAAATATGAAGAC 1536
Db 500 AsnseratgserylylstrpbylAlaAlasasnspAlaspcysgllylslslyasp 519
OY 1537 TATTAACCTCATGCGGTGACACCTCAATATATCACTGTCTTATAGTGAATGAAGAA 1596
Db 520 Tyrasprproglun-----LysThrThrllleglunlileuThrnglyAspThrarg 536
OY 1597 CAAGGTGATATACACAATAATTTTGTGACAGCTCACT----- 1644
Db 537 LysSerAspMetValglunlystrlylslysphepcysrtnrseralrthrlyAlatrghly 556
OY 1645 -----AATTAACAA----- 1653
Db 557 AlaProasprthrAlatrghlyglunsnnglyLysLysgllyAlaserglylsasnngly 576
OY 1654 GATTAATAATTAACAAAATGGAATGCTATTTTAAGATGAAAAAT----- 1698
Db 577 AspAsnllrthrghluthrthrprthrprthrprthrprthrprthrprthrprthrprthr 1698
OY 1699 -----ATTAATAGATGTAACCTGTAACACAAATATGTAATCAATCAATATATCT 1749
Db 597 LysLysAspIleasnphcysValleuunlaspolylysgln---HisThrlysgln 615
OY 1750 AAGATATATATCAATTTCAATATATTTTGAATATGCTTACATATATTAAGGATACT 1809
Db 616 LysValThrSerTyAsnValPhepetherprlystrpvallyrAspMetlellehiss 635
OY 1810 ATTAAGTGAATGACAAACTTAAACTGTATTAATATATACA---ACCACGATTTATTT 1866
Db 636 llellystrpArgasngluleuargsercyslleasnAlaLysSerGlnasnrcyslys 655
OY 1867 GAT-----GAATGTAACGAAATGCTTATGTTTTCAGACAGTGGTTAAACAAAGAA 1920
Db 656 AsnAsnlylsCysasnSerAspCysgllycysphelysglutrrpilleglunlysar 675
OY 1921 GAAGAATGATATGATTAACAAACTGTC---ACAAAAAAGAAATATACAGCAATCG 1977
Db 676 LysglutrrpAspLyslleLysAspHisphelysThrGlnsprhegllylleproglly 695
OY 1978 TAT 2037
Db 696 Proleugllylnpheasp-----TyrasrphelvalleuLysphevalleu 710
OY 2038 GACAAAGATGACGAAATGGAAGAACTATGGAATAATTAATAAAGAAAAAATGAG 2097
Db 711 Asplys-----LysgluleuendlnasnlllelylsAsprtnHisAlaasp 725
OY 2098 TTTTCAATTTGAAATAATATAGGACTATTTGACAAATGCA----- 2139
Db 726 AlaLysaspIleagllyAlaelygllyLysHisasnthrlyslleasplyspleu 745
OY 2140 -----ATGAACTCTGTTA 2154
Db 746 SerlygllygllyAsnnglyAlaLysgllyLysHisasnthrlyslleasplyspleu 765
OY 2155 GATCACTTAAGAAAGAACTGCCAGCATATGTAAGACATATATCAAAAGCAAGCATGTGAA 2214
Db 766 Glnlglunlglunlglunlglunlglunlglunlglunlglunlglunlglunlglun 783
OY 2215 ACATCCCATATGACAAACAAACCCGTGTGTTAACTCGTGAGGCAACCACT 2274
Db 784 LyslysprthrlysrvalAlaAsnproCysTy-----GlysnasnthrTyraap 800
OY 2275 AAAAAATTAAGAAATAGCACAATCTTTAAAGAGTGCATATGAGAAAGCAAGAAAT 2334
Db 801 AlaLeuAlaeglulysValAlaAlaInlleuenglnlyglunlAlaInlthrghlunleu 820
OY 2335 ---CGTGTCTTATATATGTAAGAAAGAAAGGACACAGAAAGTATATTAACGTGGGGT 2391
Db 821 HisArgserlyserleuLysgllyasnproglunlgllyHisArgsnlyasnngly 840
OY 2392 AGGAAAGACACTTCAAGACATTTATATGATATATATATATATATATATATATATAT 2451
Db 841 AsnlysserValleuLysasp---ValCyslyslleThrAsnclnlyrserlnAlaely 859
OY 2452 CTTGGTTTTCAATGACCATGTGATGCAAGCAAGCAAGGATGATGATATCAACAAAGAA 2511
Db 860 LysAsnlysserasnspProcysasnnglyLysasn-----Glnasnarg 874
OY 2512 TTTGTGTAGTGAATGAAAGGAAAGTGAATCGGAACACATGCGTAAAGATACAGAAAT 2571
Db 2512 ----- 2571

Db 875 pheasnlllelglylulstrpbyasnlglyglvalvallymetserasphrhlsser 894
QY 2572 GTTATTATCCCTCAGAGACACATATATGTATACATCCAAATTGGACATTACAAAGC 2631
Db 895 ---TyrMetProArgArgIuHISpHeCysThrSerAsnleucluhISpHeCys
QY 2632 GATGATCACCCCTTAATGTGTAATTTGTTGATTTGTTAATTCCTTTTGGG 2691
Db 914 AspLysgllygllyarphelugluvalProaspasnllyalathrlsSerleuengly 933
QY 2692 GATGTTCTCTTCATACGAAATATATGACCAACAAAGTATATACGATATATAGAAAG 2751
Db 934 AspValleuilealalalyslysgluvalagluyllelysgluuylulysas 953
QY 2752 AATAACCTAAAGGCCCAAGAAAGTAACTGACCCAAACACAGACAACTATCTGCA 2811
Db 954 LysaspGlnserGlyLysasnllalysasnlglyleuasnlglylulysThrValCysatg 973
QY 2812 GGTATATCCTTACAGTTTTCACATATAGTATATATATTCGAGAAAGATCTTGGA 2871
Db 974 AlailearglyserPhehlaspilleglyaspilleglysglyThrAspleuTrpasp 993
QY 2872 AGAAGCGGTGACATGGTAAAGCTGCAAGACATTTGGAACCTGTTTGTATATACAT 2931
Db 994 GlnasnlygllylulThrAspThrGlnserAsnleuValThrillehelglyllystle 1012
QY 2932 AAGTCATCAAGCAAGCAAGAAATGAT-----AATATATATGATGATGATGAT 2985
Db 1013 -----LysgllyThrleuasnaspThrSerlystlyAsnaspily-----LysHIS 1027
QY 2986 TTTAAATGAGGAAATGTTGGGAGCTATATAGCCAAAGTATGGAAGCCATGGA 3045
Db 1028 LeuGluLeuarglyLyspThrGluAlaasnArgAlaLysAlaTrpGluAlaMetLys 1047
QY 3046 TGTGATATATATATATGAGATTAATCGGACACCAATCAACAAAGTATATTCG 3105
Db 1048 CysAlaThrLysAspIleProaspMetLys----- 1057
QY 3106 GGATATAGTATCATACACCATGATGATATATATCCCAAAATTAAGATGATGAC 3165
Db 1058 -----CysasnlglylProIleGlnaspThrIleProGlnArgleuArgTrpMetThr 1075
QY 3166 GAATGGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3225
Db 1076 GluTrpAlaGluTrpPheCysLysAlaGlnSerGlnGluTrpAspLysleuGluTrpGln 1095
QY 3226 TGTAAAGAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3285
Db 1096 CysGlylIleCysLysLysLys-----ValGlnGlyCysThr-----SerGlyaspGly 1111
QY 3286 ---TGTACGAGTGCACAGACCTTGTAAATGATATATGATATATGATGATGATGAT 3342
Db 1112 LysCysGluLysGluLysAlaLysGluAlaCysGluAlaLysThrLysLysAspThrTrpLys 1131
QY 3343 GAACATGATGAT 3402
Db 1132 GlnGluTrpAspLysMetGluLeuLysTrpLleLeuLeuTrpAlaAsnAlaLysThrThr 1151
QY 3403 GTTATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3462
Db 1152 SerThrAsnAlaGlyllyarGhrValleuGlyAspAlaSerProaspThr-----GlnGlnMet 1170
QY 3463 ATTTGAATTTTGTGCGAATATATACAAACAA-----AATGCTGCGCAAA 3504
Db 1171 LeuAspPhePheLysLysleuGlnLysGluIleLysAsnSerAlaLeuAsnArgThrLys 1190
QY 3505 AGTAAATAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3564
Db 1191 ArgSerIleAspGlnLysAsnAspProIlePhe-----ThrSerProLysSer 1207
QY 3565 AATGTTGAGCATATCTCATGATATACAGAAATTTGATGATGATGATGATGATGATGAT 3624
Db 1208 SerAlaGluGlyllyTrlIleHISGlnGluLeuProHISThrGlnCysAspAlaGlnLysTrp 1227

QY 3625 TTTTGTGATGAAAAAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 3684
Db 1228 PheCysAsnThrAsnLys-----AsnLysGluLysTrpAlaPheArgHISGlnPro 1244
QY 3685 CAGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3732
Db 1245 HIsAspTrpAspGluAlaCysAlaCysArgProProSerThrProLysSerProAlaArg 1264
QY 3733 -----AGGTACAGATTAATAACGAAA----- 3753
Db 1265 AsnLeuProAlaAlaAspLysLysValGluGluAlaGluLysGluLysGlu 1284
QY 3754 -----AAAAACGGAGAAAGAT 3774
Db 1285 GluGluAspValGluLysAspAspAspAspAspAspAspAspAspAspAspAsp 1304
QY 3775 ACGGATGATTAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3834
Db 1305 ProGlnCysAlaLysThrValAsnAspIleLeuSerThrAspAspArgThrLysGlnValGly 1324
QY 3835 GATTTGCATCCAAAA---AAGATAGTAAT---GATATCCGATGCGAATGCGGAAAT 3888
Db 1325 GlnCysAsnAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1344
QY 3889 ATTAATTTAGTGAGACCTT---CGTGTGTATGCCCCCTTGAAGACAAAGTTATG 3945
Db 1345 SerLysPheGluAsnAsnGluTrpGlyProCysMetProProArgTrpGlnLysLeuCys 1364
QY 3946 GTACATTTCTG-----GCAATGATTAAGTAATAATAATAATTAACAATACAA 3993
Db 1365 LeuTrpTrpLeuLysGluLysGluLysGluLysAspGluLys----- 1379
QY 3994 GTTAAATTAAGAACCTTTCATCAATCTGCAGCAGCAAAACATTTCTCATGAT 4053
Db 1380 -----PheLysAspAlaLysLysLysLysLysLysLysLysLysLysLysLys 1397
QY 4054 TATTTAAAGAATGAGATGAGTGAAGAAATGAATGATGAT---AAGAATTAAGAAAG 4110
Db 1398 TyrTrpLysSerLys-----AsnSerMetAspIleLysLysLysGlnSerGly 1413
QY 4111 AAAATCTCCCGCATTTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 4170
Db 1414 GluIleProGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1433
QY 4171 TTTATGGAACAGATATATCAAAAGCTCATGCGAGGAGTAACTAAAGAGCAATA 4230
Db 1434 CysLeuAsnThrAspIleSerLysLysGluLysAlaValSerAspAlaLysGlyLys 1453
QY 4231 GATTCCTTTTCAAAAAGGTGACCAAAAATCTCTAATGGAAGAAACGCCAAGATG 4290
Db 1454 AspAlaLysPheAsnLysTrpThrAspThrAsn-----ArgThrLysTrp 1468
QY 4291 TGGACAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4341
Db 1469 TrpAspThrAsnLysProGluIleTrpGluLysMetLeuCysAlaLeuThrHISGlyVal 1488
QY 4342 AAAATGCGGCAAAAAGATGATTTTACCGAAACATGAGTGAACAAACGTCAAATTT 4401
Db 1489 ThrAsnThrAspAsnLysArgLysLysLysLysLysLysLysLysLysLysLys 1508
QY 4402 AGTGACAAAAGCACCATTTGAGAGATTTGCCAAACGACCCAGTTTTCAGATGCTA 4461
Db 1509 LysAsnValThrThrProLeuGlnLysPheAlaGluArgProGlnPheLeuArgTrpPhe 1528
QY 4462 ACCGATGAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4521
Db 1529 IleGluTrpGlyAspAspPheCysArgGluGlnLysLysLysTrpAsnGluLysGln 1548
QY 4522 AATGTAAG-----TCAATGACCAATTAAGATGATGAT--- 4554
Db 1549 LysCysAsnLysCysCysAsnAsnGlnAsnValThrSerAspLysLysLysThrLysCys 1568

QY 4555 ACAGATGTAATAGAAATGCGAGACTACGTAATATATAGAAAAAGAGTGG 4614
Db 1569 ValIuGcGlnLysCysGluLutLysGlyPhele-----ThGluTTP 1585
QY 4615 ATTCACAGATTAATATATACAGATGAAACGACAAAAAGATTGATAGAACAC 4674
Db 1586 -----GlnGluAsnTTPAsnLys-----GlnLysAsnLysTyrGluThrLeuTyr 1600
QY 4675 ATTGGTGAATGTTTACAGACTACTGAAACGAATGCAAGATTCCTGTAACAGAGAA 4734
Db 1601 ThrGlnValLysSerThrSerArgSerThrLleSerSerSerAspProIleGluThrLys 1620
QY 4735 TTTACTCTGTTGTGTGTGTAAGCTGGAAGTCCCTGTGTGTACAAAGAAATATACAA 4794
Db 1621 Leu----- 1621
QY 4795 TTGTTAGAAAAACAGCTTACTATGATGCGACAAACATTTGTGGTGCACAAATTTAT 4854
Db 1622 -----LeuLysTyrLeu 1625
QY 4855 -----GAAATATACACAAATATACTAATTCGATTAAGTAAAGTCAAGGATTAGTA 4911
Db 1626 AsnGlnLeuLysAspProLysSerAsnLleTyrSerLysAlaIleGlyTyrIle 1645
QY 4912 AAGGAGGCAAAACACAGGTGCTATTAGTGCGCAAAAGGCTCTAATATACATTAAC 4971
Db 1646 LysLys----- 1647
QY 4972 TTGAAAGAAATGACATGAGATGTCTTTTCCTTCCTGCTGACACTACGATATGTTTCAT 5031
Db 1647 ----- 1647
QY 5032 GCATGATGCGCAATATATACATTCACAGATTAAAGATGAAATGCGTGGCAAAAAACA 5091
Db 1647 ----- 1647
QY 5092 TTGATGGAAGTGGCGCAACGAGAGGTCAATTTGGTCAATACTACAAAGAAAAAAA 5151
Db 1648 -----GluGlyTyr----- 1650
QY 5152 GAAAAAGAAAAATAAAAAGTGGATGCGCAAAATATCTTATAGAGGTCCCGCTTGT 5211
Db 1651 ----- 1651
QY 5212 AGTGTATGAATATAGTTTATGATTTAAGATATATATCTAGAGTATGATTAATTG 5271
Db 1655 AsnValSerLysGlnAsnAsnPheAsp----- 1663
QY 5272 GAAGATGAAAAACAAACGAGGAGAAAAATTTGAAGAAATATTTACAAAAATGAGACA 5331
Db 1663 ----- 1663
QY 5332 TCAGTTGGCAAAAGAGATAGTACTACAGAAATCCCGGTAGTACTGCGCAAAATTT 5391
Db 1663 ----- 1663
QY 5392 TTCTGGAAGAAATTAAGCAATGTGTGTGAACCAATGATATCGGGTACAAACGTGT 5451
Db 1664 -----GluAsnLys----- 1666
QY 5452 AGGATGATGAAATAGTGAATAGTGAAGAAAGTGAAGATCTAATAAATGTGTGT 5511
Db 1667 -----AsnGlyGlyAsn----- 1670
QY 5512 TCTGTACTTCAGATGATGATATCTATGGGAAAAATCGCATGAAGTACTGCTAT 5571
Db 1671 -----AspGluLysTyrAlaPhe 1676
QY 5572 CAGTTTCTGCAGGTTTGGCGAATGGGTGAAGATTTTGAACAATAAAGAAAGAA 5631
Db 1677 -----LysGluProPolysGlu 1682
QY 5632 TTGGAAGAAATGTAAGGCGGTGAATGATTAATCTTGTGTATATAT----- 5679

Db 1683 TyrGluLys-----AlacCysLys-----CysAsnGlnAsnThrProProSer 1696
QY 5680 -----GAAGATATAAAGAAATATGACAGATCGCTGACAAATAT 5721
Db 1697 ProProGluLeuProGlyProProAlaThrAspThrSerValAspValCysGluThrVal 1716
QY 5722 AAAAAATTTATAGTAGTGAAGAACACAGATGAAAAAACAATCAAAAAATATGTGTAG 5781
Db 1717 LysSerAlaLeuThr----- 1721
QY 5782 AATTAAGCAAAATATATTCGACATCTCTGTGCAAAAGATGCAGAGCGCTCGCAA 5841
Db 1721 ----- 1721
QY 5842 TATTAAGCAAAATTAATAAATTTGTGAATAAAGTGGAGAT-----TGT 5892
Db 1722 -----IleGluAsnLeuLysGluAlaCysProThrLysTyrGlySerLysAlaProThr 1739
QY 5893 GAATATAGTGTATGAAAGATGTGTCCACACAGCGATTACTGATGTATATGCAAAAT 5952
Db 1740 SerThrLysCysIlePro-----SerGlyLysSerGlyAlaGlyAlaThrThrGlu 1757
QY 5953 ATCCCGCATCATATGACGATGAAACCAAGAAAGATTGAGAAAGTGAATGTCAAATG 6012
Db 1758 ArgSerArgAspAlaAspGlyAlaProSerLysSerGlyLysGlySerLleCysVal 1777
QY 6013 CCACAGGTCACACAGGTGATGCAAGGGAACACCGTCACACCGGATCACTGATATCA 6072
Db 1778 -----ProProArgArgArgArgLeuTyrValGlyLysLeuGluGlnTTPAlaGlu 1794
QY 6073 AAAGGAGCGATGGAATAAAGCAAGCAAGCAAGCGCCGCTACAAACCGCAAAAAA 6132
Db 1795 LysThrGlnSerSerGlnAlaGlyLysSerSerAlaThrThrGlnGluGlnValLys 1814
QY 6133 GTGAAATCTTACMACAGAAATGCGACCAAAACAGCAACCGCAGCAGCAGCAACAA 6192
Db 1815 ValAsnAsn-----LysAspGlyAlaSerSerSerSerSerSerAsnValGlnThr 1833
QY 6193 ACAGAAACCAACATCAACAGCAACAGCAACAGATCTGACGTGGCACAATGTGAAG 6252
Db 1834 ThrLeuAsnAlaLalaSerThrThrThrGluSer----- 1846
QY 6253 GCCATCTTTCGATTAACAGATAGCAGGCGGTGAATAGAGGTGTATCCAAAAAG 6312
Db 1846 ----- 1846
QY 6313 TATGCAATATCTCTAATATGGGTGTATGTAGTACTTAAGAAAAATGAAATGCG 6372
Db 1846 ----- 1846
QY 6373 ATATGATGCTCTGTAGAGAAAAAATTAATGATTAATATATATTAATTAAT 6432
Db 1846 ----- 1846
QY 6433 GAACACTGAATAAGCGTGACATGATATAAAGAGCTTTATTAATATGACAGACATA 6492
Db 1847 -----SerGlnLeuArgGlnAlaPheIleGlnSerAlaIle 1860
QY 6493 GAACACTATTTTGTGTTAAATAT----- 6519
Db 1861 GluThrPhePheLeuThrPheLysLysLysGluLysGluLysGluLysGluLysGlu 1880
QY 6519 ----- 6519
Db 1881 ArgLysLysLysGluGlnHisGlySerThrPheGlyThrLeuGluGlyLeuSerVal 1900
QY 6520 -----ATAATGAAATCTCTACACAGAAATGAATGCAAAATGCAATTCAGAT 6573
Db 1901 AspGlyGluGluGluGluGlnProProGlnGluGlnGlnSerGlyLysLleProLys 1920
QY 6574 GAATTTAAAGATATATGATATATATGATATATATATATATATATATATATATAT 6633

Db 1921 GlypheLeuArgGlmethPheTyrThrLeuAlaAspTyrLysAspIleLeuTyrSerGly 1940
Qy 6634 GATATTCTTAATGATATAAAATATA----- 6660
Db 1941 SerAsnAspAsnLeuLysIleValLeuGluAlaSerGlyThrLysGluGlySerGly 1960
Qy 6661 -----ACGTATACAAATAGTGTACACACCATCTGTCAATGCAAAATATAAGAAAAA 6711
Db 1961 AspMetGlnLysIleGlnAspLysIleLysLysThrLeuAsnGlyAspAsnAsnGlnGlu 1980
Qy 6712 CAG-----GATAAAAAAAAGATGAAGAAATACGTAATATTTGG 6753
Db 1981 SerGlyHisProSerProSerProSerGlyLysAsnSerValThrThrProGlnThrPTr 2000
Qy 6754 GAGAAATATAAAATTTATTTGGAGAGATGATATGATGATTAATCTATCTACACA 6813
Db 2001 AsnAlaHisAlaProSerIleThrIleGlyMetIleArgAlaLeuThrHisLysGlnSer 2020
Qy 6814 GAC----- 6816
Db 2021 AspAlaThrIleAlaGlyLysLysIleGluGlnAsnThrGlnLeuLysAsnAlaLeu 2040
Qy 6817 -----GAAAGCAAGAAAGAAAGAAATTAAGATTAATACAGATACATGACACCAACTG 6873
Db 2041 LeuAspAspThrLysAsnLysProLysAspAsnTyrLysTyr---AspLysValLysLeu 2059
Qy 6874 -----ACGCTTCCTCT 6885
Db 2060 AspGluAsnSerGlyThrSerProLysProAlaGlyValAsnGluAlaProProLysLeu 2079
Qy 6886 GAAAGATTTGTAAAAAGCCCAATTTTGACATGCTTACACAGATGCGCAGAAATTT 6945
Db 2080 ThrGluPheValGluThrProProLysPheArgThrLeuGlnLysThrGlyLysPhe 2099
Qy 6946 TGTAAATAGAGCAAGAAAGTGTAAATTTGAGCGGCTGTAG----- 6993
Db 2100 CysArgLysGlnLysHisLysLeuGluIleIleArgValAspCysArgGlyLysGlu 2119
Qy 6994 GAATATGAGTGAATGTAAT----- 7017
Db 2120 AspLysHisCysSerGlyTyrGlyGluAsnCysAspAspAsnLeuIleAlaAspProSer 2139
Qy 7018 -----GACGTAAGACACAGAAATGTGCGAGCGCTGTACATATCAAAATTTT 7068
Db 2140 IlePheProAspLeuAsnCysProGlyCysAlaLysHisCysSerSerTyrLysLysTrp 2159
Qy 7069 ATTAAAGAGTGGAAATGATATGAAAGCAAGCAAGAAAGTCAAAAGATTAAGAT 7128
Db 2160 IleGlnArgLysLysIleGlnPheThrGlnGlnAspAsnAlaLysTrpAsn 2176
Qy 7129 GGCAGAAAGTAAAGATTTATCTCTACTGAAAGACATGAGAGCGCAACATGTGCT 7188
Db 2177 -----GlnLysValAsnCysGlu 2182
Qy 7189 CAGTAATTTTAAACATGAATTAAGAATTTGTGCAATAAGATTTCTTTGTATG 7248
Db 2183 LysGlu-----SerLysGlyLysGlnLysValLysGlyLysLeu 2196
Qy 7249 CAAGAACTCTCTACACACTACCAAAACAAACAAATCAACATATCGATGTAAT 7308
Db 2197 Glu----- 2197
Qy 7309 GATATGCCAGATTCGCTGATTAATGTTCTGAAAGAAATTAACAAGTGTGATGCTCTGAA 7368
Db 2197 ----- 2197
Qy 7369 CTTTCAAAAAGGATCTATGATTCATACAAAAAAATTAATGACCTAAATACGTATG 7428
Db 2197 ----- 2197
Qy 7429 AATGTGTAGAGAAACAGCATATTAATTAATCAAGAA-----GCAGAAATTAAT 7479
Db 2198 -----GluAsnAlaLysPheLeuGlnLysLeuGlnLysLeuGlnLysSerLysAsnAsn 2214

Qy 7480 ATGAT-----ATTACCTTGAGCAAGAAATTTAAATTAATGCTATGAGCTTACA 7524
Db 2215 LysAspAsnGlyAspGlyThrIleAsnPheLysGlnProAspValThrPheLys----- 2232
Qy 7525 AAGCAAAAGAAAGTAAAAATAGTTGACATTAATAATACCTTCGATCTCAAGAACCT 7584
Db 2233 -----ProLysAsnLys----- 2238
Qy 7585 TATGACCTGATTAATATATAGAGAGAAACCTTGCGAAATATGAGAGAAATCGT 7644
Db 2239 -----ProCys-----SerGlu 2242
Qy 7645 TTTAAGCTGATATGATGAATGAAATGTTCACAAAATTCAAAGTCTATACAGCAAAAA 7704
Db 2243 PheLysIleAsp-----CysThrLysAlaLysCysThrGlyAspGlnGlyLys 2258
Qy 7705 AGAGTATGTGTACCTCCCAAGAGAAACATATGCTGTAAAGAAATTTAGATGAATTA 7764
Db 2259 ArgLysCys-----AsnValLysAsnGlnThrValIleArg 2270
Qy 7765 ATGAAAGACTTAAGATAGTAATTAATCTCTAAATAGTTGCTGCAACTGCACGAAAT 7824
Db 2271 AlaThrAspIleLysAspAspLys-----AsnGlyAsn 2281
Qy 7825 GAAAGATAGACATTAATAAAAAATTCAACTCAGAGACGGTGGCAATGAATCAATA 7884
Db 2282 GluAsnIleAsnMetVal-----Val 2288
Qy 7885 TGTGATACATGAATTAATTAATGTTGCGTATCTGCGTACGACATTAATGAGAACAGATATG 7944
Db 2289 SerAspThrSerLys-----LysGlyAsp-----GlnAspAsp 2299
Qy 7945 TTACGAATGCTGCTTACTTACCTCCCGTAGAATTAATTAATTAAGTTTGAATAC 8004
Db 2300 LeuArgVal----- 2302
Qy 8005 ATATATGCAAAATGAGAAATTAAGTAGAATTAATTAACACATGTACAAACG 8064
Db 2302 ----- 2302
Qy 8065 TTTGCTTCCTGCTGGCGCATGCTATAGAAAGATTTTGGAAACATGACGTGCAAA 8124
Db 2303 -----CysArg 2304
Qy 8125 GCACAGAGATGCAAAACTTTTAAAGAAAGAGAGATGATGATTTGAACGATACACA 8184
Db 2305 -----AspAlaGlyIlePheLysGlyIleArgLysAspGlnTrp----- 2317
Qy 8185 TTTAATACAGATTAAGTGTGACATTAAGACGATCCACTGTT----- 8226
Db 2318 IleCysGlyAsnValCysGlyTyrAsnValCysLysArgValLysValAsnGlnSer 2337
Qy 8227 -----GATGATTAATTAATCTCAACGTTTCGATGAGTCAAGTGTGATTAATTA 8280
Db 2338 GlyAspGlyAsnGlnIleIleIleLysAlaPheLysIleThrLeuAlaLysPhe 2357
Qy 8281 TGTAAAGCACTGATGAGAAATTTGAAGAAATTTAAATATCATGATGATCACTGT---AAA 8337
Db 2358 -----LeuGlnAspTyrAsnLysIleLysLysLysLeuLysSerGlnLys 2373
Qy 8338 ACATGTGACAGATGCAAGATGATTAATGATGAATTAAGTGTCAACGTCTAAACAGACA 8397
Db 2374 SerSerAsp-----AlaThrProCysIleLysGly 2383
Qy 8398 TGT-----CAAGATATTAATAATTTTGTCTTAATGAGAAATCTGATTCGATATACAA 8451
Db 2384 CysAlaLeuGlnThrLysLysLysLysThrThrGlnThrLysAsnLysAsnLysTyr 2403
Qy 8452 TCAATTAATAC-----AAAGAAATGTATGACACACCAATTAATTAACA 8493
Db 2404 LeuGlnGlnTyrGluAsnLysSerSerAspLysSerPheLeuValLysThrIleLeuGlu 2423

QY 8494 AAATCTCTACTTATGATCATCTTCAAAATTTTGTGTA-----CAAAAGTTGAAACT 8544
Db 2424 GlnpHeLysAspAtpProGlnpHeGlnAsnAlaIleLysProCysLysGlnLeuGlnLys 2443
QY 8545 TTTAAAGTGAATGTTCT----- 8562
Db 2444 PheGlnSerPheCysGlyLeuAsnGlyAlaAspAsnSerLysSerLysAspGlyLys 2463
QY 8563 -----GTTGAGAGCTTTTCTGAAATATCTTCATGAAACAACT 8598
Db 2464 GlnAtpAspLeuValLeuCysLeuIleGlnLysLeuGlnLys-----LysAlaGln 2480
QY 8599 AAGTTTGTGAATTTAAATTAATTAATGAATGAGTTCTTCCAAATATACGAACATATGCT 8658
Db 2481 LysCysAla-----GlnAsnAspAlaGlnAsnGlyAlaGln----- 2492
QY 8659 TTGCAAGAAACACCAAAAGTTATTAAGAAAGCTTGACATGCTTACCTCTTAAGAAAT 8718
Db 2493 -----SerCysThrGlnThrThrThrAsp 2500
8719 CCATGTGATTAATGTCTACCGACCAACAAAGAT 8754
2501 Pro-----ProThrLeuGlnAspGlnAsp 2508

RESULT 12

Q8T6L0 PRELIMINARY; PRT; 510 AA.
AC Q8T6L0;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT Identification of a conserved Plasmodium falciparum var gene
RT implicated in malaria in pregnancy."
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL; AF458886; AAL84269.1; --
FT NON_TER 1 510
SQ SEQUENCE 510 AA; 60713 MW; E3CF55DC04601099 CRC64;

Alignment Scores:

ed. No.: 3.65e-134 Length: 510
Percent Similarity: 2817.00 Matches: 501
Percent Similarity: 98.02% Conservative: 4
Query Match: 14.48% Mismatches: 5
DB: 5 Gaps: 0

US-10-087-013-1 (1-10628) x Q8T6L0 (1-510)

QY 697 AAGAAATACATGATGATGGAAGATGAAGTAAATATGATTAACATCTGATGATCT 756
Db 1 LysLysIleHsAspGlyMetGlnAspGlnValLysAsnAspLysProAsnProAspGlySer 20
QY 757 GGAATTTTAAATTAAGAGAGCATGCGATGATGATGATGATGATGATGATGATGATGATGAT 816
Db 21 GlnAsnLysLysLeuAlaGlnAlaLysLysLysLysLysLysLysLysLysLysLysLys 40
QY 817 GCATTAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 876
Db 41 AlaIleThrCysAspAlaSerLysLysSerLysLysLysLysLysLysLysLysLysLysLys 60
QY 877 CCATTAATTTTAAATCCTTAATGCGCATTAACAAAGAAAGTTCTTACCAATTTAGAT 936
Db 61 ProLeuPheSerAsnProLysCysGlyHisLysGlnGlnLysValProThrAsnLeuAsp 80

QY 937 TATGCTCCATTAATTTAGCTGTTGTCAGCAATGGGAGAGAGTTTCCGAAAAAGA 996
Db 81 TyrValProGlnThrLeuThrPheAspGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
QY 997 AATATTAAATGAAAAAGCTCAAGAGCTCTGCTAATGACAAAGAGCGTTATATGCT 1056
Db 101 AsnIleLysLeuLysLysValLysAspSerCysAlaGlnAspLysGlnLeuLysLys 120
QY 1057 AGCATTAATGAGCAATGATGTCAGCAACTAATTTGAAAAAGGATTTTGCATTGAT 1116
Db 121 SerHisAspGlyLysAspCysThrThrThrThrThrThrThrThrThrThrThrThrThrThr 140
QY 1117 AATAGTGTACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1176
Db 141 AsnLysCysThrAspCysLeuThrLysCysLysValPheGlnValThrLeuGlnLysGln 160
QY 1177 CAAGAGCATTTAAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 1236
Db 161 GlnGlnAlaPheLysLysGlnLysGlnLysLysLysLysLysLysLysLysLysLysLysLys 180
QY 1237 AACGATTAACAAATTTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1296
Db 181 AsnAspAlaLysPheValAsnAsnIleAsnSerGlnLysLysLysLysLysLysLysLysLys 200
QY 1297 CTTAAGAAACGCAATATGCAACTAATGACACTTTTAAATTAATTAATTAATTAATTAATTAATTAAT 1356
Db 201 LeuLysGlnThrGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 220
QY 1357 TATGTAAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1416
Db 221 TyrCysLysGlyGlyLeuProGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 240
QY 1417 AAGGATTAATTTTATCGTTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1476
Db 241 LysGlyIlePheThrLysSerGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 260
QY 1477 GATGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1536
Db 261 AspGlyIleLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 280
QY 1537 TATTAACCTCATGAGGTGAGAGCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1596
Db 281 TyrLysProProLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 300
QY 1597 CAAGGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1656
Db 301 GlnGlnLysPheLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 320
QY 1657 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1716
Db 321 LysAsnAsnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 340
QY 1717 GAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1776
Db 341 GlnGlnAsnThrGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 360
QY 1777 GAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1836
Db 361 GlnLeuThrPheLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 380
QY 1837 TGTATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1896
Db 381 CysIleAsnAsnThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 400
QY 1897 GACAGATGGCTTAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1956
Db 401 AspArgThrValLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 420
QY 1957 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2016
Db 421 LysLysAsnArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 440
QY 2017 TTTAAAGTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2076

```

|||||
Db 441 PheIysVal.MechrSpLysLeuAspLysGluAlaLysTrpLysLLeuMetGluAsn 460
Oy 2077 AAAAAAAGAAAAAAGAGTTTCCATTGGAATAATAGGACATATTAGAGAAAT 2136
Db 461 IleYsARgLyLysAsnGluPheSerAsnLeuGluAsnArgAspTrpLeuGluAsn 480
Oy 2137 GCATAGAACTCTGTGAATCACTTAAAAAGAACGCCAGATANGTAAGCAATAAT 2196
Db 481 AlaIleGluLeuLeuLeuAspHisLeuLysGluThrAlaThrIleCysLysAspAsn 500
Oy 2197 ACAAAAGCAAGCATGTGAACATCCCATTAAT 2226
Db 501 ThrAsnGluAlaCysGluThrSerHisAsn 510

RESULT 13
P90580 PRELIMINARY; PRT: 2647 AA:
P90580:
01-MAY-1997 (TREMBLrel. 03, Created)
01-MAY-1997 (TREMBLrel. 03, last sequence update)
01-DEC-2001 (TREMBLrel. 19, last annotation update)
FCR3-VART11-1 protein (Fragment).
DE FCR3-VART11-1.
GN Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MEDLINE=97154495; PubMed=9001213;
RA Hernandez-Rivas R., Mattei D., Sterkers Y., Peterson D.S.,
RT Wellens T.E., Scherf A.;
RT "Expressed var genes are found in Plasmodium falciparum subtelomeric
regions.";
RL Mol. Cell. Biol. 17:604-611(1997).
DR EMBL: U67959; AAC47438.1;
DR InterPro: IPR004258; PFEMP.
DR Pfam: PF03011; PFEMP. 2.
FT NON_TER 2647 2647
SQ SEQUENCE 2647 AA; 303263 MW; 65FD70611348BC7 CRC64;

Alignment Scores:
Pred. No.: 1,22e-129 Length: 2647
Score: 2727.00 Matches: 922
Percent Similarity: 34.82% Conservative: 399
Best Local Similarity: 24.30% Mismatches: 810
Very Match: 14.02% Indels: 1664
Gaps: 129

US-10-087-013-1 (1-10628) x P90580 (1-2647)
Oy 91 AGTCAAAAGTCGCAAGAAATGTTTGGACGTTATGCCAAAATATTAAGACATCATCA 150
Db 9 ThrAsnLysSerLysGluValLeuAspGluIleGlyGluThrIleGln 25
Oy 151 AAATATGCAAAAGACATGTGATTCG-----TTGAAAGGGGATTTG 192
Db 26 -----LysLysAlaHisSerSerAlaAspThrPheArgSerGlnLeuLysGlyAsnPh 43
Oy 193 ACGAAAGCAAGATTCGTGTGTGT-----CCTTCTACGCCAGTAATTAAG 237
Db 44 GlyGluAlaLysPheThrAsnGlyGlyGluIleMetGlnProAsnSerLysLys 61
Oy 238 CATAAATTATTTATTCATATTCATTAATTTAGATCATTAAGACATATAATTAACG 297
Db 62 -----CysGluLeuAspHisThrIleAspThrAsnValThr 73
Oy 298 TATGATGATGTGAATTTGAGACACTTCCTTGCCATGTAGAGACAAAACCGATTTGATGA 357
Db 74 AspGlyHisSerAsn-----ProCysGluGlyArgGlnThrValArgPheProAsp 90
Oy 358 GATGAAGAATTCGATGT---GGAATTAATAATGTAATATATAAAGAAAAAATGATGCT 414

```

```

|||
Db 91 AspAsnArgSerGlnCysThrLysAsnArgGlyLeuAsp---SerValAspAsnSerVal 109
Oy 415 ATAGCCCTGTCCGCCACACATGAGACACATATGTGATTAATTAATGCAAGCTCAAT 474
Db 110 GlyAlaCysAlaLeuProThrArgThrArgGluHisLeuCysSerHisAsnLeuGlnSerIleGln 129
Oy 475 GATATTAAT-----ACCAAAATATTCATGATTTATGGAAATGACTAGTTACA 525
Db 130 ThrAsnAspThrLysPheSerSerLysAlaLysHisAsnLeuValGluValCysTyra 149
Oy 526 GCAAAATAGCAAGGTGAATCAATTTGTAATTAATCAATCA-----CATAAAGCACT 576
Db 150 AlaLysPheGluGlyGluSerIleValLysAsnTrpGluGlnLeuGlyHisThrThr 169
Oy 577 TCAGACGCTTGACTCTGTCGACGACGTTTGCACATATAGATATGATATTAAGACA 636
Db 170 GluGlyIleCysThrAlaLeuAlaArgSerPheAlaAspIleLysPheIleIleArgGly 189
Oy 637 ATAGATATGTTTAACCAATGTCATGAC-----AAAGTAAGAACGGGT 681
Db 190 LysAspLeuThrLeuGlnLysAsnProGlnGluSerAlaArgTrpLysGlnLeuGluAspAsn 209
Oy 682 CTCGAGAGGTTTCAAGAAATACAT----- 708
Db 210 LeuArgLysIlePheGluLysLysLysGluLeuThrSerSerArgAsnGlyLysThr 229
Oy 709 GATGCAATGGAAGATGAAGAAAAAATGATTAACATCTGATGATCTGCAATTTATTAAT 768
Db 230 AsnGlyAlaGluGluLys-----AspLysSerGlyAsnTrpTrp 244
Oy 769 AAATTAAGAGACATGTGTGAATGTGAATGAATAATGAATGAAGACGTATACATGT 828
Db 245 LysLeuArgGluAspTrpAsnAlaAsnArgLeuAspIleThrLysAlaMetIleCys 264
Oy 829 GATGCAATCAATATAATCTGATTTTATTCATCAATCAAGAAATATACCATATTTTCA 888
Db 265 LysAlaProGlyLysAsnAlaProTrpPheArgAsnThrCysSerAsnGlyLysProThr 284
Oy 889 AATCTTAATATGCGCCATAAACAGGAAGAGTTTCTACCAATTTAGTATGTCCGTCA 948
Db 285 GlyLysLysCysGlnCysLysIleAspGlyThrValProThrAsnLeuAspTrpValProGln 304
Oy 949 TATTTACGTTGGTTCGACGAATGGGAGAAAGACTTTTCCGAAAAAATAATTAATTG 1008
Db 305 TyrLeuArgTrpPheGluGluTrpAlaGluGluPheCysArgLysArgAsnLeuLys 324
Oy 1009 AAAAGGTCAAGACCTCTGTCT-----AATGCAAAAAGACGTTATTTGATGT 1059
Db 325 GlnAsnAlaIleLysAsnCysAspArgGlyMetAspAspAspLysGluLysTrpCysSer 344
Oy 1060 CATTAATGACATGATTTGACACACATATTTGCAAAAAAAGATTTTTCATTGCAATTAAT 1119
Db 345 ArgAsnGlyTrpAspCysThrLysThrIleArgSerIleAspLysTrpSerMetAsnArg 364
Oy 1120 AAGTGTACTGACTGTTCGATAAATGCAAAAGTTTGAAGTTGGTAGGAAATCAACA 1179
Db 365 GluCysThrLysCysLeuTrpValCysAspProTrpValLysTrpIleAspAsnLys 384
Oy 1180 GAAGCAATTTAAAAAAGAAAAAATAATGAAAAAATAATACATATTTATTTAGAAC 1239
Db 385 LysGluPheGluLysGlnLysLysLysCysGlnAsnGluIleLysArgAsnAsnGluSer 404
Oy 1240 GATTAACAAATTTGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1299
Db 405 SerGlnAsnSerProLysAsnTrpAsnAsnMetCysGluThrAspThrArgLysAsnLeu 424
Oy 1300 AAGCAAAACCAATATGCACTAATGACACTTTTAAATTTACTAAATGAAGAAAGTAT 1359
Db 425 LysLysAspTrpGlnSerMetAsnAsp---PheLeuLysLeuLeuAsnSerGluThrPro 443
Oy 1360 TGTAAAGAGGATTTACAGAGAAAGATATTAATTTCTTAACAGTGTGTGATGACAA 1419

```


Db 444 CysThrAsnIleIleAspAlaLysSerLysIleAspPheThrLys-----AspProGlu 461
QY 1420 GGGATATTTTATCGTGCAGATATTTCCCAAGTGTGCCAGCTGGGGGCAAA----- 1473
Db 462 GluThrPheSerHisThrGlnIuTyCysAspProCysProTyGlyLeuLysThrGln 481
QY 1474 TGTATGTGTAATTAATACACACACAAATAGATATGTCGTAACGTGTAAATATGAA 1533
Db 482 AlaAspGly---ThrTrpLysArgLeuTyrgLusAsnAspProGlnCysProIleLysPro 500
QY 1534 GACATATTAACCTCCATGGGTGAGGTGACATATATCATCTGCTTATAGGTGAT 1593
Db 501 LysTrpGluProProLysGlyValGluProThrGluThrAspValLeuTyThrGlyLys 520
QY 1594 GAACAGGTGATATTCACAAATAATAGAAATTTTGTAAACAGCTCAACATATTCAAA 1653
Db 521 GluAsnLysAspIleIleValLysLeuArgGluPheCysLysThrAspGlyAsnThrGly 540
QY 1654 GATTAATAATCAAAATAAGGATGCTATTTATAGATGAAATATTAATAGATGTA 1713
Db 541 PheLys---AsnGluGluTrpAsnCysTyTrpGlnValGlyAsn---AspLysCysVal 558
QY 1714 CTGCAACAAATATCTGAATCAATTAATGATATCTAAGATATATCATTTCTATTTT 1773
Db 559 LeuGluAsnGluGluGluLeuGlyGlyLys---LysValLysAspTyTrpAsnPhe 577
QY 1774 TTTGATATATGCGTTACATATTTTATTAAGGATATATTAAGTGAATGACAACTTAA 1833
Db 578 LeuMetPheTrpValAlaHisMetLeuLysAspSerIleGluTrpTrpSerLysLeuSer 597
QY 1834 ACTTGTATTAATATCAACACACGATTTGATGATGATGATGATGATGATGATGATG 1893
Db 598 AsnCysLeuLysSerAspLysTyThrCysIleThrCysHisAspAsnCysGlnCys 617
QY 1894 TTTGACAGATGGTTTAAACAAAGAAGAAGATGATGATGATGATGATGATGATGATG 1953
Db 618 TyAspLysTrpIleGlyLysLysValHisTrpThrGlnIleLysLysHisPheAsp 637
QY 1954 AAAAAAATAATATACAG---CAATCGTATTAATGATGATGATGATGATGATGATGATG 2010
Db 638 LysGlnThrAspPheGlnIleTyTrpGlyHisTyTrpValLeuGlnThrValLeuGlnGly 657
QY 2011 ---TATTTTAAAGTATGATTAAGTGAACAAAGATGATGATGATGATGATGATGATG 2067
Db 658 AspGlnPhePheThrAspIleThrLysAlaTyrgLysAlaArgGluIleValHisIle 677
QY 2068 ATGGAATAATATAAAGAAAAAATGCTTTTCCATTTTGAATTAATTAAGGACATAT 2127
Db 678 GlnGluMetLeuGlnLysLysGluGlnVal-----LeuHisGluAspAlaSerAsn 695
QY 2128 TTAGAGAAATGCAATGACACTTTGTAGTACCTTAAAGAAAGAACAGCAATATG--- 2184
Db 696 MetLysThrIleIleAspGluLeuLeuAspHisGluLeuLysGluAlaLysGlnCysIle 715
QY 2185 -----AAGACATATAT-----ACAAAGCAAGCATGTGAACA 2217
Db 716 ValAsnHisLysAspAsnAsnCysProAlaAspLeuSerAspSerGluGlnGlu 735
QY 2218 TCCCATATGCAACAAACCCGTTGTAAACCTGCTGAGGACGCAACCCACATAA 2277
Db 736 AspIleProGlnArgGlnAsnLysCysAlaLysPro---SerGlyThrHisIleArgAla 754
QY 2278 AATTAATAAGAAATGCAACATATCTTAAAGAGTGCATGACGACGACGACGATCGT 2337
Db 755 LeuValAsnLysValAlaSerAsnMetHisLysLysLysAspGlnLeuValAsnArg 774
QY 2338 GGTCT---CATTAATGAAAGAAAGCAGCAGAGTATATTAACGTTGGGGGTAGG 2394
Db 775 GlyValSerLysLysLysGlyAspAlaLysGlyGlyTyArgLysSerGlyThr 794
QY 2395 AGAAGGACCTTCAAGCAATTTATGATTAATGATTAACATCTGATATCTGATATCTT 2454
Db 795 ThrIleLysLeuLysAsp---IleCysSerIleThrAspAspHisSerAsnAlaLysArg 813

QY 2455 GGTTTTCAATGACCATATGTATGCAAAAGCAGCATGATGATATCAAAAGATTT 2514
Db 814 GlnHisThrAspIleProCysLysArgLysAspSerLysValAsnValLysAsnArg--- 832
QY 2515 GTGCTGAAAGTGAAGGAAAGTGGATCGGAAACACATCGTAAAGATCAAGATGTT 2574
Db 833 -----ArgTrpMetAspThrAlaGlyPheIleSerAsnThrTyTrpLysAspIle 848
QY 2575 ATTATGCTCTAGAAAGCAGACATATATGATCATGCAATTTGGAACATTAACAGGAT 2634
Db 849 TyMetProProArgArgGlnHisPheCysThrSerAsnLeuGlnTyTrpLeuGlnThr 868
QY 2635 GATCAACCCCTTAATGATATATGTTGAT-----GATTAGTAAATATCTTT 2685
Db 869 AsnLysLeuLeuAsnGlnLysAsnAspIleAsnGlyAsnProAsnIleIleAsnAspPhe 888
QY 2686 TTGGGGATGTTCTTCTATCAGCAAAATATAGAAACAAAGATTAATGATGATGATA 2745
Db 889 LeuGluAspValLeuPheAlaAlaAsnTyrgLysAlaAspPheIleLysLeuMetTyAsn 908
QY 2746 GAAAGAAATTAACCTTAAGGGCCCAAGAAAGATTAAGTCAACCAACAGACATATG 2805
Db 909 LysGlnAsn-----AspTyTrpLysAspAsnAlaThrIle 919
QY 2806 TGTGAGGTATACGTGATGATTTGCAATATAGTGTATATATCGAGAGA----- 2859
Db 920 CysArgAlaMetLysTySerPheAlaAspLeuLysAspIleIleGlnArgGlnHisIle 939
QY 2860 -----GATCTGTGGAAGAAAGAGGTGACATG 2886
Db 940 CysArgIleMetIleValGluArgValLysHisGlnIleSerGluAlaGlyAsnPheLeuIle 959
QY 2887 GTAAGCGTCAAGGACATTTGGAACCTTTTGTGTAATATACATTAAGTCACTCAAGGC 2946
Db 960 LeuSerLysLysAsnIleLeu-----AlaPheLysGluIle----- 971
QY 2947 AAGGAAATGATTAATATATGATGATGATGATGATGATGATGATGATGATGATGATG 3006
Db 972 -----TyTrpGluAspThrPro---TyThrLysLeuAspGluAspTrp 985
QY 3007 TGGGAGCTATATACCAACCAAGTATGGAAGCCATGAATGTGATTAATTAATTTGAG 3066
Db 986 TrpGluAlaAsnArgLysLysIleTrpGluAlaMetGlnCys----- 999
QY 3067 GATTAATCGGACACCAATACACAAAGTATGTCGATATGATGATGATGATGATGATG 3126
Db 1000 -----ProThrProAsnGlySerPheProCysLysSerTyTrpHisIleGly 1014
QY 3127 TTGATGATTAATATCCCAAAATTAATGATGATGATGATGATGATGATGATGATGATG 3186
Db 1015 LeuAspAspTyTrpIleProGlnArgLeuArgTyMetThrGlnIleTrpAlaGluTrpPheCys 1034
QY 3187 AAGGTGCAAAAAGAGTATGATTAAGTATGAGAGCAAGTGAAGAGTGAAGATGAAG 3246
Db 1035 LysGlnGlnLysLysGlnTyrgLysLeuValSerAlaSerAsnGlnLysLysAsp--- 1053
QY 3247 GATTAATGTCAGGCTGTACGAAAGAGAGTGTACAGGTGTGCAAGTGCAGAG----- 3301
Db 1054 GluArg-ValLysValValArgIleArgIleArgValHisAsnValGlnArgAlaCysLysHisVa 1073
QY 3302 -AAGCTGTATTAATATGATTAATGATTAATGATTAATGAAAGAACATGGAATTAATA 3360
Db 1073 LysIleIleLysAsnLeuLeu---IleHisGly-----LysGlnGlnIleTrpAspLysMet 1090
QY 3361 TCAATTAATATCAAAAGATTAATGATGAAAGCAAAATGCTGTTAGTATAGTGTAT 3420
Db 1091 GluIleLysTyTrpLysLeuLeuTyTrpGlnAlaGlnThrAlaAlaAsnGlyLysPro 1110
QY 3421 GAAGCTTCAAGTACTGCCAAATATCATATATGACAGAGATGATTAATGATTTTGTGCGAA 3480
Db 1111 AspThrTyTrpSerGlyLeuValAspGluAsnGlnLysProValValAsnPheLeuPheGlu 1130

Db 1578 -----AspLysValIlyLeuaspLysn 1585
QY 5302 TTGAGAAAATATTTAAACAAATGGAACATCA-----CTGGGCAAGAAAGT 5349
Db 1586 -----SerGlyThrSerProIlyValProAlaProLys 1598
QY 5350 GATAGACTACAGGAATCCGGTAGTACTGCGCAAAATTTTCTGAGCAAGAAATAG 5409
Db 1599 ProThrThrPheProProProSerProThrSerPhe----- 1612
QY 5410 GAATGTGTGGAGCAATGATATGCGGTACAAACGTGGTAGGATGATGAAATAGT 5469
Db 1612 ----- 1612
QY 5470 GGAATAGTGCAGAAAGATGATGAAAGTCTAAAAAATGCTGCTGACCTTGCATGAT 5529
Db 1613 -----SerArgPro----- 1615
QY 5530 GATATCTCTATGGGAAAAATCCGATGAGTACTGCGTATCCAGTTCTTCGATGGTTT 5589
Db 1616 -----ProTyrPheArgTyrLeu 1621
QY 5590 GCCGAATGGGTGAGATTTTTCGAACATTAAGAAAGAAATGGAATGGAAATGGTAGG 5649
Db 1622 GluGluTyrPalaGluThrPheCysArgGluArgLysArgLeuGluLysIle----- 1639
QY 5650 GCGGTATGATTTACTTCTGCTGATGATGAAATGAAAGAAATGATCA----- 5703
Db 1640 -----LysValGluCysMetLaspGluaspLysLysGluLysCysSerGlyAsp 1656
QY 5704 ---GATCGGTGTACCAATTAATAAATTTATTAGTACGTGGAACACAGTATGAAAA 5760
Db 1657 GlyLysPaspCysGluGluIleArgLys----- 1665
QY 5761 CAATCAAAAAATTTGGTGAATTAAGACAAATATATTCGAGACCTCTGGCAAAA 5820
Db 1666 -----GlnAspLysSerThrValArgasp---PheTyrCys---ProGluCysGlyLysT 1682
QY 5821 GATGCAGAGACGCTCCGCAATTTTATAGACAAACAAATTAATAATTTGTGAATAA 5880
Db 1682 Yrcys-----ArgPheTyrLysArgTyrPileGlyLys----- 1692
QY 5881 AGTGGAGATTGCAATATTAAGTGTATGAAGATGTGTCCACACAGCGATTAATCTGATGT 5940
Db 1692 ----- 1692
QY 5941 AATAGTCAAAATATGCCCGCATCATTAAGACATGAACAAAGAAAGTGAAGAGTGT 6000
Db 1692 ----- 1692
QY 6001 AATGTGCAAGTGCACAGAGGTCCACCAGTGTAGAGAGAAACCGTCACACGGGT 6060
Db 1692 ----- 1692
QY 6061 TCACGTATATCAAAAGCGACGCGCATGAAAGAAAGCAAGCAAGCCCGCTTCAAAA 6120
Db 1693 -----LysLysaspGluTyrAspLysGlnLysGlu-----AlaTyrAsn 1706
QY 6121 CAGCCGAAAAAGTGAATATCAACAGAAATGCGAGACAAACAGAACCCGAGCA 6180
Db 1706 snGlnLysThrAspAlaArgAsnAsn-----AsnAspAsnAlaPhe 1721
QY 6181 GCAGCACAACAACAGCAAAAGCAATCAACAGACACACACAGAAATCTGAGCTGGG 6240
Db 1721 eThrThrLeuaspThrCysThr---ThrAla-----Gly 1731
QY 6241 ACAATGTAAAGCCATCTTTGCAATPAACACAGATAGCAGGGGTGAATAGAGGTTGT 6300
Db 1732 AspPheLeuGlnThrLeu----- 1737
QY 6301 AATCCAAAAAGTATGACAAATATCTTAATGGGTTGATAGTAACTTAAGAA 6360
Db 1737 ----- 1737

QY 6361 AATGAAATGCAATATGATGCTCTAGAGAAAAAATATATGATTAATATACAA 6420
Db 1738 ---LysAsnLysProCys----- 1742
QY 6421 TATTTAAATTAATGAAACGTAAGAAATGACGATCAATATTAAGAGCCTTTATTAA 6480
Db 1743 -----LysAsnAspAsn----- 1746
QY 6481 TGTGACCAATAGAAATCAATTTTGTGTAAATATATTAATGAAATCCTGACCA 6540
Db 1746 ----- 1746
QY 6541 GAAATGAATTCGAAATATGCAACAAATTCAGATGATTTAAAGAAATATATAC 6600
Db 1746 ----- 1746
QY 6601 TATGTGATTTAATGAATATGTTTGTGAACTGATATTTCAATGATAAAAATATA 6660
Db 1747 -----ValAspAspSerGlyGluAsnLys----- 1754
QY 6661 ACTGTAAACAATAGTAAACAACATCTCAATGAAATTAATAGAAAAAACAGATAA 6720
Db 1754 ----- 1754
QY 6721 AAAAAAGATGAAGATTTACGTAATATTTGGAGAAAAATTAATTTGGGAA 6780
Db 1755 -----LysIlePhe----- 1757
QY 6781 GSAATGATATATGATTAATTAATTAATCACTCAACAGCAAGAAAGAAAAAATTA 6840
Db 1758 -----AspGluAsn-----Gly 1761
QY 6841 GATATTTACGATCAATGACATGACCAAACTGACGCTTCCCTTGAAGATTGTAAA 6900
Db 1762 AspThrPheLysTyrThr----- 1767
QY 6901 AGCCCCCAATTTTGAGATGTTCAAGAAATGGCAGAAAGATTTGTAATTAAGAGAC 6960
Db 1768 -----GlnTyrCysGly----- 1771
QY 6961 CAACAGTTGTAATTAATGAGCGCGCTGTAAGAAATAGTATGATAGTACTAATGAC 7020
Db 1772 -----ThrCysSerLeuAsnGlyPheLysCysAsnLysasp--- 1784
QY 7021 GGTAAACACAAGAAATGCGAGAGCGGTGTAAATATTAATTAAGAGTGG 7080
Db 1785 -----CysArgValArgThrAsn----- 1790
QY 7081 AAAAATAATATGAAGACAAAGAAAGAAAGTCAAAAAGATTAAGATGCAAAAAGTAT 7140
Db 1790 ----- 1790
QY 7141 AAGGATTAATCTTCTACTGAAAGACATTAAGACAGCAACATGCTCATGATATTTA 7200
Db 1791 -----ValThrCys----- 1793
QY 7201 AACATGAATTAAGAAATATGATGCAATTAAGATTTCTTGTATGCAAAAACCTTCT 7260
Db 1793 ----- 1793
QY 7261 TCACACTACAAAAACAACAACAATCAATCAATCGATCGATGATGCAAGAA 7320
Db 1794 -----AsnGlySerAsnArgThrThrIleThrAlaAspIle----- 1807
QY 7321 TCCTGTGATTAATGCTCAAGAAATTAACAGATGAGTGCCTGAACCTTCAAAAAG 7380
Db 1808 -----LysAsn 1809
QY 7381 GATCTATGATTCATCAAAAAAATTAAGTAAACCTAAATACATGATGATGTAGAG 7440
Db 1810 Gly----- 1810

QY 7441 AAGACGATATTTATTTATCTTAAGACGAGAAAATATATG-----GATATT 7488
Db 1811 -----GlySerSerAlaGluIleAsnMetLeuValSerAspIle 1824
QY 7489 ACCGTGAAGCAAAAATTTTACTTATGAGTCTACAAAGAAAAGAAAGTAAATATAG 7548
Db 1825 AsnSerGlyAsnGlyPheAsnAspLeuGluVal----- 1835
QY 7549 TGGACTAATTAATTAATCCCTTGCATCCTTAAGAAAACCTTATGCACGTAAATATATAGGA 7608
Db 1835 ----- 1835
QY 7609 AGAAGAAACCCCTTGTGAATAAGAGAAAATCGTTTAAAG-----GTAGATTATGAA 7662
Db 1836 -----CysLysAsn-----AlaAsnIlePheGlyLysGluAsnLys 1849
QY 7663 TGGAAATGTTAACAAAATTTCAAGTCTATCAAGAGAAAAGATATGTACCTCA 7722
Db 1850 TTrpLysCys-----ValTyrPheCysLysSerAspValCys----- 1861
QY 7723 AGAAGAGAACATATGTGTTAAGGAATTTAGATGAATTAATAAGACTTAAGAT 7782
Db 1861 ----- 1861
QY 7783 AGTAAATATCTCTAATAATGTTGTCGACATGCGACGAATGAGGAATAGCATATAA 7842
Db 1861 ----- 1861
QY 7843 AAAAATCTCACTCAGAGAACGGTGCAGATGATCAATATGATATCTATGAATAT 7902
Db 1861 ----- 1861
QY 7903 AGTTTCGTGATCTGGTGACATAGTTAGAGAACAGATATGTTACGAATTTGGTTAC 7962
Db 1861 ----- 1861
QY 7963 TTAACCTCCGTAGAAATTAATATATTAAGTTTGAATACATATATGAAAATGAGAA 8022
Db 1861 ----- 1861
QY 8023 AATTAATAATAAGTAGAATAATTAATCAACAGATGACAAAACGTTCTGCTGTGTG 8082
Db 1862 -----GlyLeuLysLysAsnAsnAspIle----- 1869
QY 8083 GATGCTAATAGAAAGATATTTGGAAGCAATGACGTGCAGAACCCAGAGATGCAAAA 8142
Db 1870 AspGlnAsnGlnIleLeuLeuIleArgAlaLeu----- 1880
QY 8143 CTTTTTAGAAAAGAGATGATGATTTGAACGCATTAATATACAGATAAGTGT 8202
Db 1881 -----PheLysArg----- 1883
QY 8203 GGACATTAAGAGATCCACCTGTGATGATTAATATCTCAACGTTTCGATGATGACT 8262
Db 1883 ----- 1883
QY 8263 GAATGCTCAATATATTTGTAAGCAGATGAGAGAAATTTGAAAATTTAAATATCA 8322
Db 1884 ---TTrpLeuGluTyrPhe-----LeuAspAspTyrAsnLysIleArgLysLys 1898
QY 8323 TGTGATCATGTAAACATCTGACAGATGCAAGATGATTAAGTAAATAAGTGTGA 8382
Db 1899 LeuAsnProCys-----IleAsnAsnGlyGluLysAla 1909
QY 8383 CAGTGTAAACGAGATGTCAGATATTAATAATTTTGTCTTAATGAGAAATCTATATC 8442
Db 1910 TLeCysThrAsnGlyCysValGlu-----GlnTrp----- 1919
QY 8443 GATATATCAATCAATTAATTAACAAGATTTGTATGAACAACATATATACAAAATCTCT 8502
Db 1919 ----- 1919
QY 8503 ACTTATGATCATGTCAAAATTTTGTACAAAAGTTGAAAACTTTAAAGTGAATGTCT 8562

Db 1920 ---IleAsnHisLysArgTyrGluTyrPheAsnLeuLysSerPhe----- 1933
QY 8563 GTTGAGAGCTTTTCTGAATATCTTACGAAAACAAGTAAAGCTTGAATTAATAATTAAT 8622
Db 1934 ----- 1934
QY 8623 GAAATATGATGCTTCTTCATATATAGAACATATATGCTTTCGAAACACCCAAAAGTTAT 8682
Db 1939 GlyAspAsp----- 1941
QY 8683 AAAGAAAGCTTGACGTTGTACACTACTCTTAAGATTCATGATTAATTTCTCCATCCGAT 8742
Db 1942 ----- 1942
QY 8743 CAAACAAAGATGATGATTAAGAAATTAACAACCTTTACCTTCTGTCGAGAAATGATAT 8802
Db 1944 ArgAsn----- 1945
QY 8803 GATTAATTAATCTTGATTAATTTGGAACGATACCTTGTCTTAATAGTTCAGATGATACAAA 8862
Db 1945 ----- 1945
QY 8863 GGTGTATGATTCCTCCAGAGAACAGACATTTATGTACAAAGACCTATACCTGCATTAAT 8922
Db 1946 -----ProArgLeuArgPhe----- 1950
QY 8923 TATAGAAAAGTGATTAAGAAATTTTAAAAAAAACCTTCTACTTCTGCTTCAAGTCAA 8982
Db 1950 ----- 1950
QY 8983 GGACAAATGTTAGTCAAAAATTAATTAATCGAAGAAAGTGTGCTTGAAGCAATGAAA 9042
Db 1950 ----- 1950
QY 9043 TATAGTTATGAGATATTTCCGATATATTAATTAAGAACTGATATGATGCACTTCAATTA 9102
Db 1951 -----PheValAspLeuIle----- 1955
QY 9103 TCTGAAAAAATTAATAATTTTGAACATCAATATGAACCAACGAAAATCTTAATAACA 9162
Db 1955 ----- 1955
QY 9163 TGGTGGAAAAATTAATAGACGTCAATATGCGACGCTATGTTATGAGATTAATAATTTGCT 9222
Db 1956 -----ArgGlnIle----- 1958
QY 9223 ACTTCAAAAGTATACATTAATTAATGAGATGCTGTCAATTAACCAAGATGAAAGAACTAAT 9282
Db 1959 -----AlaAlaThrIleAspLysGly-----Asn 1966
QY 9283 CAGTTCTCTGCTGTTAATTTGAATGGCCAAAGCAAGCATGTAAGCAAAAAGAAACATGTA 9342
Db 1967 His-----AsnGlyLeuValLysLeuValLys----- 1975
QY 9343 AGTATTCATTAATAACAAATGCTCCTCGTTCAACAGAGATATTTTGAAGCGTCAGAA 9402
Db 1976 ---SerValLysCysAsnCysGlyAsnAsnSerGlnAsnGlyLysGluGlu 1993
QY 9403 TTAATTAAGACACCTGATGCTCAGAAATGATTAAGAAATATATTTACCTTGAATATATG 9462
Db 1994 ----- 1994
QY 9463 ATTAATAATTAATCAATGAAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9516
Db 1997 ValLeuCysLeuLeuGlnLysLeuGluLysLysAlaGluLysCysLysAspAsnProGlu 2016
QY 9517 TCTTCAGGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9576
Db 2017 ThrSerGly-----IleProGlnGlnProCysGluValSer-----ProAsnHisIleGluAsp 2034
QY 9577 AAAGATTCATCAATGCGCTTGGAGTTAAATGATATAT-----AATGAA 9618

```

Db 2035 GIUGLUGLProLeuGIUGLUGLusnThrValGIUGLisProLysIleCysaspasp 2054
QY 9619 AATAGTACAGAAACAAATAATGAATTAATGATTCACAAAGAGTACTATAAAATTA 9678
Db 2055 ValLeu-----LysHisasnHisasnGlnArgasnGlnGlnArgLeuValLys-As 2071
QY 9679 TATCGGTGTTATTTGTTGTAAGATGAACACACAAAAATCATAGTACTAGTGAAT 9738
Db 2071 nProLeuValGlnProThrLeuLysArgLysLysLysLysLysLysLysLysLysL 2091
QY 9739 AATAAAGAGAGACAAACAGTTCGTTAAGACAGCTATTTCTTACACCCAGCTA 9798
Db 2091 e-----LysLysLysasnGlnAsnProHisProArgHis-Leu-----ProCysG 2106
QY 9799 GATTTCTTCTCAACAGCACTTTATTTCTCAACACATGAGTACATATGATCT-- 9855
Db 2106 LysAlaPheLeuasnThr-----AsnThrProLysThr-LysThrProProSer 2121
QY 9856 -----AAAAATGAT-----ATATTGAAAAGTACTATCTCTGTTGTTAT 9894
Db 2122 SerGIYusasnProThrPrgLInHisProAlaValIleProAlaLeuValThrSerThrLeu 2141
9895 GTATCGCGCTAGGTGTTG-----ATAGCGCTCATTTTCATGAGAAATAATTC 9942
Db 2142 AlaTrpSerValGIYIleGlyPheAlaAlaPheThrTyrrPheTyrrLeuLysLysLysThr 2161
QY 9943 AAATCGTCTGTGACGTCTGTCCGATCTGATCTGAATATCCCGACAGAGATGATGATGCT 10002
Db 2162 LysSerThrIleAspLeuLeuLeuSerLeu-----IleProLysSerAspLysPhePro 2180
QY 10003 ACCTGGAATCCAAATAAGTACATACATATGAGATGCTCCATTAAGGCAAAACA 10062
Db 2181 ThrLysLeuSerProAsnArgTyrrIleProTyrrThrSerGIYusTyrrArgLysLysArg 2200
QY 10063 TATATATATATGAGAGAGTACT-----AGTGGAGATGAAGATAAATATATGATG 10113
Db 2201 TyrrIleTyrrLeuGlnGlnLysAspSerGIYThrAspSerGIYThrAspHisTyrr----- 2218
QY 10114 GACTATATCTCTGCTGATATTTACT--TCATCCGAAAGAGATGATGAGATGATAT 10170
Db 2219 -----SerAspIleThrSerSerSerGIYusGlnIleGlnIleMetAspIle 2234
QY 10171 AATGATATATATGATACAGTACTGCTAAATATATAAACAATGTAGTAGTACTAGTA 10230
Db 2235 AsnAspIleTyrrValProGIYSerProLysTyrrLysThrLeuIleGlnValLeuGln 2254
QY 10231 CCATCA----- 10236
Db 2255 ProSerGIYusLeuSerGIYasnThrIleProThrSerGIYasnAsnThrThrAlaSer 2274
10237 -----AAAAGGATATACCAAGTATGATACACCAAGTAAATGATACACACGTACGAAT 10290
Db 2275 AspThrGlnAsnAspIleProThrSerAspThrProProIleThr----- 2290
QY 10291 AGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10350
Db 2291 -----AspAspIleThrAsnThrLeuLysHisAspPheIleSerAsnMetLeuGln 2307
QY 10351 AATACAGAACCAAT-----AATAATACAAAGATGCAATATTCATGATACAGAA 10404
Db 2308 AsnGlnProLysAspValProAsnAspTyrrThrSerGIYasnSerSerThrAsnThrAsn 2327
QY 10405 CCTAATACCTTATATTCGAT----- 10425
Db 2328 IleThrThrThrSerArgAspAsnValAspAsnAsnThrHisProThrMetSerArgHis 2347
QY 10426 AATCTGAGAGAAAACTTTATATATATATATATATATATATATATATATATATATAT 10485
Db 2348 AsnValAspGlnLysProPheIleThrSerIleHisAspArgAsnLeuTyrrThrGlyGln 2367
QY 10486 GAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10545
Db 2368 GlnTyrrAsnTyrrAsnValAsnMet-----ValAsnThrMetAspAspIleProIleAsnSer 2386

```

```

QY 10546 AGAATGATCTTATAGAGTATAGATTAATTAATTAATTAATTAATTAATTAAT 10604
Db 2387 HisAsnValIleTyrrSerGIYIleAspPheIleAsnAspThrLeuSerGIYasnGlnHis 2406
QY 10605 ATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10628
Db 2407 IleAspIleTyrrAspGlnLeuLeu 2414

RESULT 14
ID Q26033 PRELIMINARY; PRT: 2664 AA.
AC Q26033:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Variant-specific surface protein.
GN VAR-2.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
RT "The large diverse gene family var encodes proteins involved in
RT cytoadherence and antigenic variation of Plasmodium falciparum-
RL infected erythrocytes";
RL Cell 82:89-100(1995).
DR EMBL, L40609; AAF5398.1;
DR InterPro: IPR004258; Pfam: PF03011; PDB: 2A64;
DR Pfam: PF03011; PDB: 2A64;
SQ SEQUENCE 2664 AA; 302410 MW; 6EA2468511703091 CRC64;

Alignment Scores:
Pred. No.: 1,29e-129 Length: 2664
Score: 2726.50 Matches: 906
Percent Similarity: 34.34% Conservative: 393
Best Local Similarity: 23.95% Mismatches: 879
Query Match: 14.01% Indels: 1605
DB: Gaps: 122

US-10-087-013-1 (1-10628) x Q26033 (1-2664)
QY 97 AAAAGTGCAGAAATGTTTGGAACTTATGCCCCAAATAATTAAGACATCCATAAATAT 156
Db 17 ArgSerIleLysHisIleLeuAspSerIleGIYLysLysVal-----Tyrr 31
QY 157 GCAAAAGACATGTGATTCG-----TTGAAGGCGATTGAGC 195
Db 32 AspLysValHisGlyAspAlaLeuGlnProSerAsnGlnLysLeuGlnIleLysLeuSer 51
QY 196 AAAGCAATTTGCGTGGTCTCTTACGCCAGTAAATGACATATTAATTAATCA 255
Db 52 LeuAlaIlePheGlnLysAlaProGlnGlnLysGlnThrSerGlnLysP----- 67
QY 256 TATCATGTAATTTAGATCAATAAGCAATACTAATTTACGG-----TATGATGATGCG 309
Db 68 ---ProCysAspLeuAsnHisGlnIleGlnIleThrValThrSerGIYTyrrAspLysGln 86
QY 310 AATTGAGACATCTTCCATGGTGAAGACAAACCGATTGATGATGATGATGATGATGAT 369
Db 87 Asn-----ProCysLysAspArgProGlnValArgPheSerTyrrThrGlnIleLysAla 103
QY 370 GAATGTGGA-----AATAAATAGTAAATTAATTAATTAATTAATTAATTAATTAAT 426
Db 104 GlnCysAspLysSerLysIleArgGlySerAsnSerAsnLysAspGly---AlaCysAla 122
QY 427 CCAACTAGAGAGACGACATATGTTGATTAATAAATCTGGAAGCTCTA---AATGATATAAT 483
Db 123 ProPheArgArgLeuHisLeuGlnHisLeuGlnHisIleLysHisAspLysIle 142

```

QY 484 ACCCAAAATATTCATGATTTATGGAAATGCTACTGATTAACAGCAAAATACGAAGTGAA 543
143 ThrArg-----HisasnLeuAlaAspValCysGluAlaIleLysPheGluAlaGlu 160
QY 544 TCAATT-----GTTAAATATCATCATCAATAAAGAACTTCA 579
161 SerLeuGluLysIleYrArgGlyGlnIYrGlnLeuAsnSer-----Ser 174
QY 580 GACGCT-----TGTAAGTCTCTGTCACAGAGTTTGCAGATATATGATATAT 627
175 AspValasnIleasnIleCysThrGlnLeuAlaArgSerPheAlaAspIleGlyAspIle 194
QY 628 GTAAAGGAATAGATATGTTAAACCAAT-----GTCCATGACAAAGTAAGAACGGT 681
195 ValArgGlyYrArgAspLeuYrArgGlyAsnAspLysGluLysAspArgLeuGluAsn 214
QY 682 CTCGAGAGGTTTCAAGAAATACATGATGAAATGGAATG---GAAATGAAATATGAT 738
215 LeuArgLysIlePheLysIleYrAspAsnLeuAsnSerAlaHisValGlnGluHis 234
QY 739 TACAAATCCGAT-----GGATCTGGAATATATATATTAAGAGAACAGATGGGAAT 792
235 TyrLysAspAspAspLysGlyThrLysAsnIleYrLysLeuArgAsnAlaIlePheGlu 254
QY 793 GTCAATAGAAATAAAGTATGGAAGCTATACATGTCATGATCATCATATAATCTGATAT 852
255 AlasAsnArgIleThrValTrpLysAlaIleThrCysGlyAlaAla-----GlyGlyThr 272
QY 853 TTTATGCAATCAGAAAGTATACACATTTATTTCAAAATCTTAATGCGGCATTAACAA 912
273 TyrPheArgGlnIleThrCysGlyThrGlyThrTrpHisAsnIleLysCysArgCysProIle 292
QY 913 GGAAGAGTTCCTACCAATTTAGATATGTCCTCAATATTTAGCTGGTTGCAAGATG 972
293 AsnAspValProThrTrpPheAspTyrValProGlnIleuArgTrpPheGlnGluIle 312
QY 973 GGAAGAGTTCCTGCGCAAAAGAAATATTAATTTGAAAAAGTCAAGAGCTCTGCTG 1032
313 AlaGlnAspPheCysArgLysLysLysLysValAspIleValLysThrAsnCysArg 332
QY 1033 AATGACAAAGACGCTATTTATGATCATATGACATGATGATGATGACAAATATTTGG 1092
333 AsnTrpSerArgAsnLeuIYrCysSerGlyAsnGlyLeuAspCysGlnIleuArg 352
QY 1093 AAAAAAGTATTTGCTATTTGATTAATAGTACTGCTGCTGCTGCTGCTGCTGCTGCT 1152
353 ValIleGlyHisHisValIleGlySerGluCysSerLysCysSerValTrpCysArgArg 372
QY 1153 TTGGAAGTTTGGTTAGGAATCAACAGACATTTAAAAAGCAAAAGAAATATGAA 1212
373 TyrLysLysTrpIleAspAsnGlnLysGlnGluPheLeuLysGlnLysLysCysGlu 392
QY 1213 AAGAAATACAAATCATATTTATGCAAGATACAAATTTGCAATATATATATAGGAA 1272
393 AsnGluLeuLeuSerLysSerLysLysGlnSerThrLysTrpAsnValTrpGlnGly 412
QY 1273 TATATTAACAATTTATGAAAACTTAAGAAAACCAATATGCAACTTAATGACACTTT 1332
413 TyrAspArgGluPheYrLysIleLeuLysSerGluAsnValGlyGlyLeuAspLysPhe 432
QY 1333 TTAATTTACTAAATGAGAAAGTATTTGAAA-----1335
433 LeuGluLeuLeuAsnGlnGluArgGluCysGlnGluPheSerAsnAspLeuGlyLysIle 452
QY 1366 -----GGAGATTCACAGAGAAAGAAAGATATTTACTTTTAAACAGT 1407
453 AsnPheLysThrValAspGlyGlyValAlaIleGlyGlyAlaSerAspSer 472
QY 1408 GCTGATGACAAAGGATATTTATCTTCAAGATATTTGCCAAGTGTCTCCGACTGCGG 1467
473 AsnAsnSerAsnLysTrpPheSerHisSerGlnIYrCysGlnGluCysProGlyCysGly 492

QY 1468 GTCAAATGTATGTTATTAATACACACACAAATCAATATATATCTGTGAACGTGAAT 1527
493 ValGluLeuIleGlyAsnGluTrp-----LysIleLysAsnLysGlyLysCysLysGly 510
QY 1528 AATGAACATATTAACCTCCATGGGTGTGAGAGCTACTATATATCACTGCTTTATAGT 1587
511 GlyAsnArgGlyTrpAsnIleProLysGlyThrLysHisAsnValIleProValLeuSerPhe 530
QY 1588 GGTAAATGACAAAGGTATTTACACAAATTTAGAAATTTTGTAAACACCTCAACTAAT 1647
531 GlyAspLysIleLysGluIleIleGlyLysIleGluLeuPheCysAlaGluSerAsnSer 550
QY 1648 TACAAAGATTAATAATCAAAATGGAATGCTATTTATTAAGATGAAATATTAATAGA 1707
551 AspSerSerLysLeuThrGlnIleTrpLysCysTrpTyrGlyAspLysGlyTrpGluVal 570
QY 1708 TGTAATCTGGAACAAATTAATGAAATCAATATGATTAATCTTAATGATA---ATATCATTT 1764
571 CysThrLeuGluAsnArgAsnLysSerGluGluAspProGluIleGlnLysThrPhe 590
QY 1765 CATTAATTTTGTGAATATGATGATTAATTTATTAAGGATACTATTAAGTGAATGAC 1824
591 HisAsnPheThrPheTrpIleArgHisLeuLeuAsnAspSerIleGluTrpArgAsp 610
QY 1825 AAATTTAAATCTGTATTAATTAATACACACAG---CATTTGATGATGATGATACAGA 1881
611 LysIleAsnAsnCysIleGlyLysAlaLysGluLysGlyLysCysLysAsnGlyLysThr 630
QY 1882 AATGCTTATGTTTGTACAGATGGTAAACAAAGAAAGAAAGAAAGTAAATGATTAAG 1941
631 AspCysGlyCysPheGlnAspTrpIleGlyLysLysLysGlnGluTrpGlyLysIleLys 650
QY 1942 AAACGTTCACAAAAAAGAAATATACACACATCTATTATTAATTAATTAATCTT 2001
651 LysHisPheLys-----ThrGlnAspGlyPheSerIlePheGlyAsn--- 664
QY 2002 TTTGAAAGTATTTTAAAGTTATGATTAATCTTACAAAGTGAAGCAAAATGGA 2061
665 -----AsnTrpAspPheValLeuGluAsnValLeuAsnIleAsp----- 677
QY 2062 GAACCTATGAAATATTAATAAAGAAATGAGTTTCCATTTGAAATTAATTAAG 2121
678 GluLeuPheGlnAspIleThrGlnAlaIleAspAspAlaThrGlnGluLysAsnThr 717
QY 2122 GACTATTTA-----GAGCAATGCA 2139
698 AsnTrpLeuAlaLysLysLysThrGlnAlaAlaAspAspAlaThrGlnGluLysAsnThr 717
QY 2140 ATAGAACTCTTTGATGATCACTTAAAGAAAGTCCACGATATGTAAGCAATTAATACA 2199
718 IleAsnLeuLeuPheGluTrpAspSerGlnGluAlaLysCysLys-----LysIle 735
QY 2200 AACGACATGTGAAACATCCAT---AATGCAACAAACAAACCGGTGTTAAACGCTG 2256
736 GlnGluLysGlnGlnProLysLysProThrLysValArgAsnProCysTrp----- 752
QY 2257 GGAGCAGCAGCAACCACTAAATATTAAGAAATACCAATCACTTAAAGAGAGTGA 2316
753 GlyAsnAsnThrTrpAspAlaLeuAlaGlyLysValAlaGlnLysLeuGlnGlnGluAla 772
QY 2317 TACAGAGACG---CGAATCGGTGCTTCATTAATTTGAAAGAAAGCAGCAAGGT 2373
773 LysGlnGluLeuAsnAspArgAsnAspSerArgSerAlaLeuLysAlaAsnAlaSerGlnGly 792
QY 2374 ATATATTAACGCTGGGTAGAGAAAGACATCAAGACATTTATGATAGATATAGATA 2433
793 LysTrpSerAsnGlnGlyAspProAspAspPheLysLysAsnLeuCysGlyIleThrGln 812
QY 2434 AAACATTTCAATCGTAATCTGTTGTTTCAATGACCAATGATGACAAAGCAGCAGT 2493
813 LysHisSerAsnAla---IleGlyAspSerLysAsnProCysAsnAsnLysGly----- 829
QY 2494 GATGCTATACAAACAAATTTGCTGTAGGAACGTAATGGAAGTGGATCCGGAACACATG 2553


```

Db 830 -----LysGluArgPheAsnValGlyLysTrpLysAsnGlyGlu-----Val 845
OY 2554 CGTAAAGATCAGCAGAGATTTATTTATGCTTCTAGAACGACATATTTGATCCAAAT 2613
Db 846 LysMetSerHisThrAspLeuTrpLeuProPheArgGlnHisPheCysHisSerAsn 865
OY 2614 TTGGACATTTACAAACGAGATCACCCACTTAATGTAATATGTGTGATTTACTT 2673
Db 866 LeuGlnHisLeuAsnThrLys-----SerThrGlyLeuThrSerAspLysAla 881
OY 2674 AATAATTCCTTTTGGGGAGTCTTCTGATACAGCAAAATTTGAAGCAAAAGATATA 2733
Db 882 IleHisSerLeuLeuGlyAspAlaLeuLeuAlaLys----- 894
OY 2734 CGAATGTATTAAGAAAGAAATACCTAAAGGCCCAAGAGTAAGTACCCAAACAC 2793
Db 895 -----LysGluGlyLysAlaPheLys-----ThrLysLeuThrGlnAsnAspAsn 909
OY 2794 CAGCAACTATCTGCGAGTATACGTTTACAGTTTGCGAGTATAGGTGATATATTCGA 2853
Db 910 ArgSerSerIleCysArgThrMetLysThrSerPheAlaAspIleGlyAspIleLeuArg 929
OY 2854 GGAAGAGATCTCTGGAAAGAAAGCGTGACATGCTAAAGCTGCAAGACATTTGGAAACT 2913
Db 930 GlyThrAspLeuTrpAspIleAsnGlyAspAlaThrGlyValGlnAsnAsnLeuLysAsp 949
OY 2914 GTTTTGGTAAATATACATAGCTCAGCAAGCGCAAGAAATGATAATATATATGAT--- 2970
Db 950 IlePheSerLysIleThrGlnGluLeuLysGlnHisProAspLysPheAsnAspAsn 969
OY 2971 -----GATGCCCAAAATATTTAAATTTGAGGAAATTTGGCGGAAGCTTAAT 3018
Db 970 AspLysTrpThrAsnAspSerLysHisThrLysLeuArgSerAspTrpGlnAlaAsn 989
OY 3019 AGAGCCAAAGTATGGAGAGCCATGAAATGTATATAAATTTGAAGATAATGCGGA 3078
Db 990 ArgAspValValTrpLysAlaMetThrCysProThrLys----- 1002
OY 3079 CACCAATCAACACAAAGTATTTGCGGATATAGTCATATACCACTTGGATGTTAT 3138
Db 1003 -----AsnGlyAsnIleGlnCysGly-----AlaThrProHisAspAspLys 1016
OY 3139 ATCCCAAAAATTAAGATGATGATGACGATGGCAGAAATGATGATGACGAGGAGAA 3198
Db 1017 IleProGlnArgLeuArgTrpMetValGluTrpAlaGluTrpPheCysLysGluInSer 1036
OY 3199 AAGGAGTATGATGATGATGAAAGAAAGTGAAGAGTGAAGATTAAGATTAAGTCA 3258
Db 1037 ArgLeuTrpGluLeuLeuLeuArgAspCysGlySerCys-----ThrThrGlyLys 1053
OY 3259 GGCTGTACGAAAGAGAGTGTACAGTGTGTACAGAGTGCACAGAACTTGTAAATAT 3318
Db 1054 -----CysAsnAsnAspLys-----CysAlaLysCysAspLysGlnCysGlnLys 1069
OY 3319 AATGATATATAGGATTTAGAAAGCAATGAAATATATATATGATATATACAAAGAA 3378
Db 1070 LysTrpLysIleGlnProTrpAlaAspGlnTrpAsnGluIleSerAsnLysTrpGlnIle 1089
OY 3379 TTACATGAAAGACAAATGCTCTGTAGTATATAGTGTATGAACTCCAGTACTGCC 3438
Db 1090 LeuTrpTrpGlnAlaLysIleAlaIleAsnGlyLysTrpGlnLysSerThrThrThr 1109
OY 3439 AAAAATCATATAGACAGATGTTATGAAATTTTGTGCGAATTTATACCAAAAGTGT 3498
Db 1110 LysAspAspLysAspLysAsnValIleAspPheLeuGlnLysLeuHisGlnAlaAsnTrp 1129
OY 3499 GGCAAAAGTATTAAGGCTACTAGTGAAGT----- 3534
Db 1130 Gly-----ThrArgGlyProProGlnAlaHisProAspArgArgProArgArg 1146
OY 3535 GCTGTCATGTGTTACTAACACACGATGAAATATGTTGAGACATATTCACCATGATACAG 3594

```

```

Db 1147 AlaAlaThrSerLysSerAspValTyGluThrThrAlaGlyTrpIleHisGlnAla 1166
OY 3595 AATTTGATGATTTGTCAGTGCAGCAAAATGAGTTTGGATGAGAAAGATGCTAAGAT 3654
Db 1167 ArgThrArgGluLysLeuGlnAsnValPheCysAsnAsnAsnGly----- 1182
OY 3655 AACGAAAATATATGCTTTAGATTAAGTAACACAGACCATGATGTCGTGTGTAA 3714
Db 1183 AsnAsnGlnLysTrpAlaPheSerLeuThrProHisGluTyLysHisAlaLysLysCys 1201
OY 3715 AGTGATCGAAACCGAACGAGGTACGATTAAGCAAAAAAGCAAGCAAGCAAGAT 3774
Db 1202 -----AsnGluAsnLysAlaSerSer----- 1208
OY 3775 ACGGAATGTAAACAGTGAATGATATCTTAAGAAACAGTGAAGAAACAGTAGAA 3834
Db 1209 -----ProGluGluLeuGlyArgSerAspSerPhe 1218
OY 3835 GATTGTCATCCAAAAAAGATAGTAAATGATATCCGATTTGGCAATGGCAATTAAT 3894
Db 1219 AspAspHis----- 1221
OY 3895 TTAGTGAAGACCCGTCGTGTATGACCCCTAGAACAAAGTATTCGTACATTTTC 3954
Db 1222 -----GlnThrProArgProGluGluAspGluValHis--- 1232
OY 3955 TTGGCAATGATTAATGAATAAATAAATAATTAACATCAAGTTAAATTAAGAGCTTTC 4014
Db 1232 ----- 1232
OY 4015 ATCAATCTGCAGCAGCAAGAAACATTTCTTCATGCTATTTATTAAGATGAGATGT 4074
Db 1233 -----SerSerGluGlnGly 1237
OY 4075 GAGGAAATGAACTCGATTAAGAAATTAAGAGCAAAATTCCTCCCATTTTGAGA 4134
Db 1238 GlnLysAspGluSerGluAspGluGluGlu----- 1248
OY 4135 TCATGTTCTACACATTTGGAGATTTATAGATTTTATTTGGAACAGATATATCAAAA 4194
Db 1248 ----- 1248
OY 4195 GGTGATGTTGAGGAGTAATACTAAAGAGCAATATGATCTCTTCAAAAATGTTGAC 4254
Db 1249 -----GluGluValGluLysValHisAspGlyAlaAsp 1259
OY 4255 CAATAATCTCTATATGAAAAAACACGCCAAGATGTGGACAGACATGATGATGATA 4314
Db 1260 GluLysAla----- 1262
OY 4315 TGGAGCTATGCTATGTGCTAGTATGTAATAATTTGGGCAAAAAAGATGATTTTACGAA 4374
Db 1262 ----- 1262
OY 4375 AACTACGTTTACAACAAGTCAAAATTTAGTACAAAAAGCACACTTTGGAGAAATTTGCC 4434
Db 1262 ----- 1262
OY 4435 AAAGACCCCGTTTTTACGATGCTAACCAGATGTGACAGCATTTGCTATACACGA 4494
Db 1262 ----- 1262
OY 4495 CAAAATATTTGAAGATGTGCAGAAAAATGTAAGTCAATAGCAAAATGAAAGTGTAT 4554
Db 1262 ----- 1262
OY 4555 ACAGAAATGTAATAGAAATGCGAGACTAGCTTAATATATGAAAAAAGAGTGG 4614
Db 1262 ----- 1262
OY 4615 ATTCCAAAGATTAATTTACAAGATGAACGCAAAAAAAGATTTGATAGACACAC 4674
Db 1262 ----- 1262

```

QY 4675 ATTGGTGAATGGTTACAGACTATCTGAGCAAGATGCAACAGATTCTTGAACAGAAA 4734
 DB 1262 ----- 1262
 QY 4735 TTCTACTGCTAGTTGGTGTAAGCTCGAGAGTGCCTGTGTGTACAAAGAAATATACAA 4794
 DB 1262 ----- 1262
 QY 4795 TTGTAGAAAAACAGCCTTACTATGATGCCGACAAACATTGGTGGCACAATAATTATT 4854
 DB 1262 ----- 1262
 QY 4855 GAAATGACGACAAATATCTAATCATTTGAGTAAAGATTAAGTCAAGAGATTAGTAAAG 4914
 DB 1262 ----- 1262
 QY 4915 GAGGCAACACAGTGCCTTATTAAGTGGCAAAACAAAGTCCATAATACATAACTTG 4974
 DB 1263 ----- 1273
 1263 -----GlyAlaValSerGlnProGluAlaSerProThr----- 1273
 4975 AAGAATTGACTGAGATGCTTTTCTCTCTGCTGACTAGTATATGTTTCATGCA 5034
 DB 1274 -----ThrLysAspValAla----- 1278
 QY 5035 TTGGATGCAATTTATACAGATCCAGAAATTAAAGATGAAATGGGTGGCAAAAGATTG 5094
 DB 1278 ----- 1278
 QY 5095 ATGGAATGCGCGCAACGAGAGGTACATTTGGTCACTACTACAAAGAAAAAGAA 5154
 DB 1278 ----- 1278
 QY 5155 AAGAAGAAATATAAACGTCGATGCGCACAATATTTCTATGAGTCCGCTGTAGT 5214
 DB 1279 -----LysProProCysAsp----- 1283
 QY 5215 GCTATGAAATATAGTTTATGATTTAAGACATATATCTAGTATATGATTAATTTGAA 5274
 DB 1284 -----GluLeuPheSerAsnValAspThrLeuGln----- 1296
 1284 -----LysProProCysAsp----- 1296
 QY 5275 GATGAAAAAAGCAAGCAGAGAAATTTGAAGAAATTTTAAACAAAATGAGAACATCA 5334
 DB 1296 ----- 1296
 QY 5335 GTTGGCAAGAGATGATGATCTACAGAGAAATCCCGTAGTACTGCGGAAAAATTTTTC 5394
 DB 1296 ----- 1296
 1296 ----- 1296
 QY 5395 TGGACGAAAAATRAAGATGTGTGTGAACGCAATGATATGCGGGTACAAACGTGTAGG 5454
 DB 1296 ----- 1296
 QY 5455 GATGATGAAAAATAGTGAATAAGTGCAGAAAGTGAAGATCTAAAAAATGTTGTC 5514
 DB 1297 -----LysAlaCysSerThr----- 1301
 QY 5515 GTACCTTCAGATGATGATTTCTCTATGAGGAAAAATCCGATGAGAGTACTGCGTATCAG 5574
 DB 1302 -----LysTyrGlyProGlyLysAsn----- 1308
 QY 5575 TTCTCTCGATGCTTTCGCAATGGGGTGAAGATTTTTCGAAACATTAAGAAAAAGAAATTG 5634
 DB 1308 ----- 1308
 QY 5635 GAGAAATGTTAGGGCGGTGTAATGATTAATCTGTGTGATTAATGAAGTAAAAAGAAAG 5694
 DB 1308 ----- 1308
 QY 5695 AATATGACAGATGCGTGTACACAATATAAAAAATTTATTAGTACAGTGAACCAACAGTAT 5754
 DB 1308 ----- 1308

QY 5755 GAAAAACAATCAAAAAATATGTTGAGAAATTAAGACAAAAATATATTCGAGCATCTGTG 5814
 DB 1309 -----TyrGly----- 1310
 QY 5815 GCAAAAATGACAGAGCAGCCTCGCAATATTTAGACAAACAATTAATAAAATTTGTGAA 5874
 DB 1310 ----- 1310
 QY 5875 AATAAAGTGAAGATTTGAAATTAAGTGAAGATGATGTCACACAGCATTAAGT 5934
 DB 1311 -----TyrArgCysIleProThrLysThrSerAsnAspValThr----- 1323
 QY 5935 GATGTAATAGTCAAAAATATGCCCCATATTAGACATGACAAAGAAAGTTGAAGG 5994
 DB 1324 -----GlyGluAspGlyGln----- 1328
 QY 5995 AAGTGAATTTGCAAGTGCACAGAGTCCACAGCGTGAAGAGGAAACACCGTCACCA 6054
 DB 1329 -----GlySerArgArgValAlaIleArgSerThrProGlu----- 1339
 QY 6055 CCGGTATCACTGATATCAAAAAGCAGCAGCATCGAAAAAGAAACAGCCGCCCT 6114
 DB 1340 -----SerGlySerAsnSerAspLysAsnGlyAlaThrCysIleProPro----- 1354
 QY 6115 ACAAAACAGCCGAAAAAGTGAATACTAACACAGAAATGCGACACAAACAGCAACC 6174
 DB 1355 ArgArgArgArgLeuTyrValGlyLysLeu----- 1364
 QY 6175 CGACGACGACACACAAACGAAACGAAACATCAACGACCAACACAGATCTGAC 6234
 DB 1365 -----GluGlnThrAlaAsnLysHisAsnThrCysIleThrSerValSerGlnGly----- 1380
 QY 6235 GTGGCACAATGTTAAAGCCATTTCTTGAATTAACACAGATGACAGGGTGAATAGAG 6294
 DB 1381 GluAlaThr-----GluAlaArgGly----- 1387
 QY 6295 GGTGTAATCCAAAAAGTATGACATATCTTAAATGGGGTGTATTTAGTAAAGTCT 6354
 DB 1388 -----SerGluAlaProAlaProGly-----Gly----- 1395
 QY 6355 AAGAAGAAATGAATGCAATATGATGCTCCCTCCTAGAGAAAAAATTAATCTATTAATAT 6414
 DB 1396 LysGluSerSerSerGly----- 1401
 QY 6415 ATCAATATTTAATATGAACTGAACCTGAATAATTAACCTGACATGATATAAAGACGCTTT 6474
 DB 1402 -----GlyLysGluThrProSerAspLys-----LeuArgThrAlaPhe----- 1414
 QY 6475 ATTAATATGTCACCAATAGAACTCAATTTTGTGTTAAATATATTAATGAA----- 6528
 DB 1415 IleGluSerAlaIleValGluThrPheLeuThrPaspArgTyrLysGluTyrPleu----- 1434
 QY 6528 ----- 6528
 DB 1435 AlaGluLysLysAlaGluLeuGlnAsnGlyLysLeuAspLeuTyrSerSerGlyAspGly 1454
 QY 6529 -----AATCTGACAGCAAGAAATGAATTCGAAATGGAACAATTCACAGATGAATTT 6579
 DB 1455 AspProAspAsnPro-----GlnAsnLysLeuLeuAsnGlyAlaIleProProAspPhe----- 1472
 QY 6580 AAAAGATAATGATATATATATATGATGTTGATTAAGATATGTTTGGAACTGATAT 6639
 DB 1473 LeuArgLeuMetPheTyrThrIleGluGlyAspTyrArgAspIleLeuValHisGlyLysAsn 1492
 QY 6640 TCTAATGATTAATAAAATTTAATCACTGAACAATAAGTGAACAACATTCATCAATGAAT 6699
 DB 1493 ThrSerAspSerGlySerThrAsnGlySerAspAsnAsnAsnIleValLeuGluAlaSer 1512
 QY 6700 AATTAAGAAAAACAGATATAAAAAAGATGAAGATTAACGTAATAA----- 6747
 DB 1513 GlyAsnLysGluAspMetGlnLysIleGlnGluLysIleGluIleLeuProLysAsn 1532
 QY 6748 -----TTTGGGAGAAA----- 6759

Db	1533	GLYGLYThrProLeuValProLysSerSerAlaGlnThrProAspLysTrpTrpSerGln	1552
QY	6760	AATAAATAATTTATTTGGAGGAGATGATATATGATTAATTAAT	6804
Db	1553	HisAlaGlnSerIleThrPheLysGlyMetIleCysAlaLeuThrTrpGluLysAsnPro	1572
QY	6805	-----CATCTCACAGACGAAAGGAAAGAAAAAATATAGATTAATTAAC-----	6849
Db	1573	AspThrSerAlaArgLysAspGluAsnLysIleGluLysAspAspGluValTyrGluLys	1592
QY	6849	-----	6849
Db	1593	PheThrSerThrIrrAlaAspLysHisGlyThrAlaSerThrProThrGlyThrTyrLys	1612
QY	6850	---CAGTACAAAT---GACATGACCCAAACTG-----	6873
Db	1613	ThrGlnTyrAspTyrGluLysValLysLeuGlnLysAspThrSerGlyAlaLysThrProSer	1632
QY	6874	-----ACGCTTCCTTGAAGAGTTTGTAAAGAGCCCAATTTTGGAGATG	6921
Db	1633	AlaSerSerAspThrProLeuSerAspPheValLeuArgProPheArgTyr	1652
QY	6922	TTTCAGCAAGATGGCAGAGAAATTTTGTAAATACAGAGAAAGAGCTGTAAATTTGGAG	6991
Db	1653	LeuGluGlnTyrPheGlyAsnAspPheCysLysGluArgLysIleGluLysGlnIleLys	1672
QY	6982	GGCGGCTGT-----AAGCAATATGAGTGAATGCT-----	7011
Db	1673	GluGluCysMetLysAspLysSerAspLysLysTyrSerGlyAspGluGlnCysAspArg	1692
QY	7012	-----AGTAATGACGGTAAAGACACAGAAAGTGGACAGCGCTGT	7050
Db	1693	ArgAspThrSerAsnGlnValSerAlaAspLeuGluLysArgSerCysGlyAsnSerCys	1712
QY	7051	GTAACATATCAAAATTTTATTAAGAAGTGGAAACGTAATATGAAGACAGAAAG	7110
Db	1713	ArgPheTyrLysLysTrpIleLysArgLysArgLysGluTyrAspLysGlnAlaAsnAla	1732
QY	7111	TTCAAAAGAGATTAAGATGGCAAAAGTATTAAGATTAATTCCTTCTACTGAAGACATA	7170
Db	1733	TyrSerLysGlnLys-----ThrLysTyrGlu-----	1741
QY	7171	GAGAAAGCAACATGTCCTCATGAAATTTAAACATGAATTAAGAATTATGTGCAAT	7230
Db	1741	-----	1741
QY	7231	AAGATTTGTTTGTATGCAAAAACCTTCTTCAACACTACCAAAAACAACACAATCA	7290
Db	1742	-----GluGlySer-----	1744
QY	7291	CAATCATCGGATGCTAATGATATGCGAGAAATCGTGATTAATGTTCTGAGAAATTTAAC	7350
Db	1745	LysGlyAlaLysLysAsnAsp-----HisAsn-----	1753
QY	7351	AAGTGTGAGTCCCGCAACTTTCAAAAAAGGATCATGATTCATACAAAAAATTTACT	7410
Db	1754	LysGluPheCysValLysLeuLys-----	1761
QY	7411	GAACCTAAATTAATCTATGATTTGTAGAGAAAGCAGCATTTAT-----	7455
Db	1762	-----ThrCysThrAspAlaIleAlaPheLeuAsnArgLeuLysAsn	1775
QY	7456	-----TTATCTAAAGAGCAAAAT---AATATGATATTATACCTTG-----AAG	7497
Db	1776	GlyProCysLysLysAspAsnGlnLysnGlyLysAsnAspIleAsnPheGlyAsnThrGlu	1795
QY	7498	GAATAATTTATACGATGATGCTACCAAGCAAGAAAGAAAGTAAATATGTTGACTAAT	7557
Db	1796	GluThrPheArgProAlaGlnAsnCysLys-----	1805
QY	7558	AATAATCTTGCGATCTCTAAGAAACCTTATGCACTGATTAATATATGGAAGAAAG	7617
Db	1806	-----ProCysSerSerPheLys-----IleAsnCysArgAsn	1810
QY	7618	---CCTGTGAAATATAGAGAAAGAAATCGTTTAAAGTAGATTAATGAAATGCAATGAC	7674
Db	1817	GlyAsnCysArgSerGlyAspGly-----AspThrLysGluLysCys-----	1830
QY	7675	AAAAATTCAAAGTTCTATCAGAGAGAAAAAGAGTATGTATACCTCCCAAGAGAGACAT	7734
Db	1830	-----	1830
QY	7735	ATGTCTTAAGGAATTTAGATGAATTAATAATGAAAGCTTAAGATGATTAATTCCTC	7794
Db	1830	-----	1830
QY	7795	CTAAATATGTTCTTCGCAACTGCACAGAAATGAGAAATGACATTAATTAATTAATTAAC	7854
Db	1831	-----	1841
QY	7855	TCAGAGAACGGGTCGCAATGATATCAATATGTATCTATGAATATAGTTTGGCTGAT	7914
Db	1842	ThrMetGlyThrCysThr-----	1847
QY	7915	CTGGTGACATAGTTTAAGAGCAAGATATGTTTCAATTTGGTGGTTACTTACCTCCGTA	7974
Db	1848	---GluAspValVal-----	1851
QY	7975	GAATAAATAATATTAAGTTTGTGATTAATATGAAATATGAAATATGAAATATTA	8034
Db	1852	-----MetHisValSerAspLysAsnAlaAsnGln	1861
QY	8035	GGTAGAAATTAATTAACAAGATGTAACAACGTTTCTGTTCTGGTGGATGCTAATAGA	8094
Db	1862	PheGluGlyAspGlyLeuAspArgLysAlaCysGluLysAsnIleGlyIlePheThrGlyIleArg	1881
QY	8095	AAAGATTTTGGAAAGCAATAGAGCTGCAAGACACAGAGATGCAAAATTTTGAAGA	8154
Db	1882	LysAspLeuTyrLys-----CysArgLysValCysGlyLeuHisIleCysLysGln	1898
QY	8155	GGAAGATGATGATGATTT-----GAACGATTAATTAATTAACAAGTAAATGTGGA	8205
Db	1899	GluLysGlyAsnGlyAlaIleAsnAspGlnGlnIleLeuValArg-----	1914
QY	8206	CATTAAGAGCATCACCCTGTTGATTAATATATACCTCAACGGTTTCAGATGACTGA	8265
Db	1915	-----AlaLeuLysArg	1919
QY	8266	TGGTCTGAATATATATGTAAGACATGATGAGAGATTTGGAATTTAAATATCATGT	8325
Db	1920	TyrValGluTyrPhe-----LeuGluAspTyrLysLysIleLysLysLysLeu	1935
QY	8326	GATCACTGTAAACATCTACAGATGCAAGATTAATGATGAATTAAGTGTGAACAG	8385
Db	1936	LysProCys-----IleGluAsnGlyAsnGlySerThr	1946
QY	8386	TGTAAACAGAGATGTCAGAAATATTAATAATTTTGTCTTAATATGAAATCTATTCAT	8445
Db	1947	CysIleAsnGlyCysAsnLysLysCysAsnArgValGlyLeuTyrIleLysLeuLysLys	1966
QY	8446	ATACATCAATTAATTAACAAGATTTAT-----	8475
Db	1967	AspGluTyrThrLysIleLysAsnHisTyrLeuGluLysAsnLysGluLysAspLysAsn	1986
QY	8476	-----GACACCAATATATTAACAATAATCTCTACTTATGAT	8511
Db	1		

OY	8633	GAATATGATGGTCTTCACATATTAACGACATATATGCTTTCGAAAGAAACCAAAAAGTTAT	8662
Db	2031	Lysmsnspolysasnclushn	2037
OY	8663	AAAGACCTTGCAGTTGTACACTACTCTCTAAGAAATGCATTGGATTTGCTACCGAT	8742
Db	2038	---AspAlaIleAspCysmelleu	2044
OY	8743	CAAAACAAAGATGGATGTAAGAAATTAACAACCTTTACTCTTGCTCCGAAGATCATAT	8802
Db	2045	-----AsnLysLeuGluThrLysIleHisGluCysLysThrGlnHis	2058
OY	8803	GATATATATCTTGATATATTTGGAACGCATCTTGTTCTTAATAGTTTCAGATGAT	8856
Db	2059	GluhsnbservAlaGluhsn	SetAspGlnProHis 2065
OY	8857	---AACAAAGGTGTATTTAGTATCTCTCCAAAGAGAGACATTTATGTACAAGACCTATAC	8913
Db	2070	ProAsnCysGlyGlyAsnProProPro	2078
OY	8914	GCATATATATTAAGAAAAAGGTGATTAAGAATTTTAAAAAACTTCTTACTTCTGCT	8973
Db	2078	-----	2078
OY	8974	TTTCAGTCAGACAAATTTAGGTGTCAAAAATATTAATCGAAGAGAGTGGCTTTGAG	9033
Db	2079	-----AspGluGluAspLeuLeuGlu	2086
OY	9034	GCAATGAATATATAGTTATGACAGATTTATCCGATATATTAAGAACTGATATGATGAC	9093
Db	2087	Glu	2087
OY	9094	ACTTCATTTACTGAAAAAATTTAAAAAATATTTGAACATCAAAATGAAGAACCCAAAT	9153
Db	2087	-----	2087
OY	9154	CGTAAACATGTGGGAAAAATTAATAGACGTCAAGATATGCGACGTATGTTATGTGATAT	9213
Db	2088	-----GluAsnProValGluGln	2093
OY	9214	AAAATTCCTACTCCAAAAGTAACATTAGATGAAGAGTGGTGCATTTACCAAAAGATGAA	9273
Db	2094	-----ProGlyPheCysProThrProGlnGlnGlu	2103
OY	9274	GAACCTATCAGTTTCTGCTGTGTTAATTAATGAATGGCAACCAACGATGAAGAAAG	9333
Db	2104	ProGluProAsp	2107
OY	9334	AAACATGTAAAGTATTCATTTAAAAACAAAATGTCTCGTTCAACGAGATATTTTGA	9393
Db	2108	-----AspLysCysGlyLysLeuGlnGluLysLysAspGlu	2119
OY	9394	GGCTCAGAAATTTATAGACAACTCGATGTCAGATGATATTAGAAAATATATATAGCTTG	9453
Db	2120	LysLysGlu	2124
OY	9454	AATATATGTATAAAAATAACAATGAAGAAATCTAAATATATAAATATTAAGCAATTAAAGAT	9513
Db	2124	-----	2124
OY	9514	CAATCTTCAGGTATATATAGCAATTAACATCTGAAGAAATGTTCAAGCATATATTA	9573
Db	2125	-----GluGlnProAlaGluGluAspGlyGlyAlaIleValPro	2137
OY	9574	TCAAAAATTTCTCAATCCGCTTTGGAGTTAAATGATATATAATGAANAATAGTTACAGAA	9633
Db	2138	SetGlyPro	2140
OY	9634	AAAAATATGAATAATATGAATCAACAAGAGTACTAAAAAAATTAATCCGTGTTATAT	9693
Db	2141	-----ProGly	2142

OY	9694	TTTGTGTAAGATGAAGAACACACAAAATATCATGACTGACTGAGAAATATTAAGAACAGAG	9753
OY	2143	---SerGIuPrOgIuAlaAspLySgLyProValIySProAlaIguIle-----	2157
OY	9754	CAAAACATGTCGCTCAAGACACTCATATTTCTTTACACCCACATGATTCCTTTATGCA	9813
Db	2158	-----ProIySPro-----	2161
OY	9814	GCACCTTATATTCACACACATGAGTACGACAAATATGATCCGAAAAATGATTAATGA	9873
Db	2162	GIuPrOProAspLeuSerHisProAla-----	2173
OY	9874	AGTACTATCTGTGTGTTATGTTATCGCGTTAGTTG-----	9921
Db	2174	SerLeuValThrSerThrLeuAlaThrSerValGlyIleGlyPheAlaAlaPheThrTy	2193
OY	9922	CATTTCATGAGAGAAAAATTCAAATCGTCGCG---GACTTGTCCGTAATCTGAATATC	9978
Db	2194	PheTyrLeuValYsLySgLyThrTySerSerValGlyAsnLeuPheGlnIleLeuGlnIle	2213
OY	9979	CCGCAAGAGAGATGATGAGATGCGCTCGTTCGTAATCCAAAAATGAGTACTACTACATAGA	10038
Db	2214	ProIySProAspTyAspLeuProThrIySLeuSerProAsnAlaTyrlleProIyThr	2233
OY	10039	AGTGTTCATATTAAGCAAAACATATATATATATATGAGAGAGATACT-----	10089
Db	2234	SerGIyLySgTyArGIyLysArTyrlleTyrlleuGIuGIyAspSerGIyThrAspSer	2253
OY	10090	GGAGATGAAGATAAATATATATGCGGACCTTATCTCTCGATATACTTCAATCCGAAGT	10149
Db	2254	GIyTyrlThrAspHisTy-----	2267
OY	10150	GAGTATGAAGATTTGATATTAATGATATATATGTCACGAGTACGCTAAATATGAACA	10209
Db	2268	GIuTyrlGIuGIuMetAspIleAsnAspIleTyrlValProGIySerProIySgTyrlSthr	2287
OY	10210	TTGATGAGAAGTGCTACTGAAACCATCAAAAAGGATATACCAAGAT-----	10257
Db	2288	LeuIleGIuValValleuGIuProSerGIyAsnAsnThrThrAlaSerGIyAsnAsnThr	2307
OY	10258	-----GATACACCAAGT-----	10269
Db	2308	ThrAlaSerGIyAsnAsnThrThrAlaSerGIyLysAsnThrProSerAspThrGlnAsn	2327
OY	10270	-----AATGATACACACCGTACGATAGATTTATGATGATGATGATGATGACTG	10320
Db	2328	AspIleGlnAsnAspGIyLleProSerSerIySgIleThrAspAsnGIuTyrlAsnGlnIleu	2347
OY	10321	AAACATGATTTGTATGCCAATATTTA-----	10371
Db	2348	IySAspGIuPheIleSerGIuIyLeuGlnSerGIuProAsnThrGIuPro-----	2365
OY	10372	TACAAAGTCAGATATTCCAATGATACAGACAACTAATACTTATTAATCTGATATCTCT	10431
Db	2366	MetLeuGIyTyAsnValAspAsnAsnThrHisProThrThrSerHis--HisAsnVal	2384
OY	10432	GAAAGAAAAACCTTTATATATATATATCTATCTATGATAGGATTTTATATACGTGGCAAGAAAT	10491
Db	2385	GIuGIuIySgProPheIleMetSerIleHisAspArgAsnLeuPheSerGIyGIuGIuTy	2404
OY	10492	AGTAT-----	10518
Db	2405	AsnTyrlAspMetPheAsnSerGIyAsnAsnProIleAsnIleSerAspSerThrAsnSer	2424
OY	10519	-----ACTAATATGATATTCATGATGATGCTAGAAATGATTCCTTATAGA	10563
Db	2425	MetAspSerLeuThrSerAsnAsnHisSerProTyrlAsnAspLyAsnAspLeuTyrlSer	2444
OY	10564	GGTATAGATTTAAATATGATGACTAGCGTGCAAAACCTTATGATATATATGATGAAG	10623
Db	2445	GIyIleAspLeuIleAsnAspAlaLeu--SerGIyAsnHisIleAspIleTyrlAspGIu	2464
OY	10624	TATTCG 10628	

```

Db      2464 etlenu 2465
RESULT 15
ID      094657      PRELIMINARY;      PRT;      2212 AA.
AC      094657:
DT      01-FEB-1997 (T-EMBLrel. 02, Created)
DT      01-FEB-1997 (T-EMBLrel. 02, Last sequence update)
DT      01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE      Erythrocyte membrane protein 1.
GN      PEPMP1.
OS      Plasmodium falciparum.
OC      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX      NCBI_TaxID=5833;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN=ITG;
RA      Yang S.O., Wollish W.S., Gut J., Wu J., Ahn J., Petersen C.,
RA      Fujioke H., Alkawa M., Leech J.H., Nelson R.G.;
RA      "The molecular cloning and DNA sequence analysis of Plasmodium
RA      falciparum erythrocyte membrane protein 1."
RA      Submitted (Jul-1995) to the EMBL/Genbank/DBJ databases.
DR      EMBL; U31083; AAB0661.1;
DR      InterPro; IPR004258; PEPMP.
DR      Pfam; PR03011; PEPMP; 2.
SQ      SEQUENCE      2212 AA;      253594 MW;      A33456648CC852PC CRC64;

Alignment Scores:
Pred. No.:      3.37e-127      Length:      2212
Score:      2678.50      Matches:      719
Percent Similarity:      44.45%      Conservative:      311
Best Local Similarity:      31.03%      Mismatches:      809
Query Match:      13.77%      Indels:      478
DB:      Gaps:      84

US-10-087-013-1 (1-10628) x 094657 (1-2212)
QY      91 AGTCACAAAGATGCAAGAAATGTTTGAAGCTTATGCCAAATATTAAGCATTCATCA 150
DB      170 GUGUgLyIleCYsthrAlaLeuAlaArgSerPheAlaAspIleArgIleArgGly 189
DB      9 ThrAnLysSerAlaLysGluValLeuAspGluIleGlyIleThrIleGln----- 25
QY      151 AAATATGCAAAAGAACATGTCGATTGCG-----TTGAAAGGGGATTGTC 192
DB      26 -----LysLysAlaIleHisSerAspAlaSerPheArgSerGlnLeuLysGlyAsnPhe 43
QY      193 ACGAAGCAAGATTCGCGGTGCT-----CCTTCAGCCCACTAATAATAG 237
DB      44 GYgluAlaLysPheTyrAsnLysGlyIleGluIleMetGlnProAsnSerLysLeu----- 61
QY      238 CATATATATATATATCATATCATATGATCATATTAAGTAAGAACATACATATTAAGC 297
DB      62 -----CysGluLeuAsnPhIsthrIleAspThrAsnValThr 73
QY      298 TATGATGATGATGATTTAGACACATCCCTTGCCATGCTAGACAACACGATTTGATGA 357
DB      74 AspGlyHisSerAsn-----ProCysGluLysArgGlnThrValArgPheProAsp 90
QY      358 GATGAAAGATCTGAATGT---GGAATATAAATACGATATTTTAAAGAAAGAAATGATGCT 414
DB      91 AspsnAsnArgSerGlnCysThrLysAsnAsnArgIleLysAsp---SerValAspAsnSerVal 109
QY      415 ATAGCCTGTGCGCCACCTAGAGAAGACGATATGTGTATAAACCTGGAAGCTCTAAT 474
DB      110 GYAlaCysAlaProTyrArgArgLysHisLeuLysSerHisAsnLeuGlnSerIleGln 129
QY      475 GATTAATAT-----ACCCAATATTTATGATATTATTGGAAGAAATGATCTAGTACA 525
DB      130 ThrAsnAsnTyrAspSerSerLysAlaLysHisAsnLeuAlaGluValCysTyrAla 149
QY      526 GCAAAATACGAAGGTGATCATCATGTGTATATATATCATCA-----CATTAAGGAAGCT 576
DB      150 AlaTyrPheGluGluGlnSerIleValLysAsnTyrGluGlnLeuGlyHisHisThrThr 169

```

```

QY      577 TCAGACGCTTGACCTCTCTTCACAGAACTTTGCAGATATAGTGATATGTAAGACA 636
DB      170 GUGUgLyIleCYsthrAlaLeuAlaArgSerPheAlaAspIleArgIleArgGly 189
QY      637 ATAGATATGTTTAAACCAAAATGTCATCCATGAC-----AAGTGAACACGGGT 681
DB      190 LysAspLeuTyrLeuGlnLysAsnProGlnGlnSerAlaArgArgLysGlnLeuLysAsn 209
QY      682 CTCGAGAGAGCTTTCAAGAAATACAT----- 708
DB      210 LeuArgLysIlePheAspPheIleTyrLysGluLeuThrSerSerArgAsnGlyLysThr 229
QY      709 GATGGAATGAGAGAGAGACTAAATATGATTCATCTGATGATCTGGAATTTATAT 768
DB      230 AsnGlyAlaGluGluArgTyrLys-----AspGlySerGlyAsnTyrTyr 244
QY      769 AAATTAAGAGAACATGTCGATGTCATATGATTAATTAAGTATGGAAGCTATTAACATGT 828
DB      245 LysLeuArgGluAspThrTrpAsnAlaAsnArgLeuAspIleTrpLysAlaMetIleCys 264
QY      829 GATGATCATATTAATCTGATATTTTATGCAATTCAGAAAGTAATACACATTAATTTCA 888
DB      265 LysAlaProGlyAsnAlaGlnTyrPheArgAsnThrCysSerAsnGlyGluLysProThr 284
QY      889 AATCCTAATGCGGCCATTAACAAAGAAAGGTTCCTACCATTTATGATATGATTCAT 948
DB      285 GlyLysLysCysGlnCysIleAspLysThrValProThrAsnLeuAspTyrValProGln 304
QY      949 TATTATGCTGTGTCGACGAATGGGGAGACAGTTTGGCCAAAGAAATATTAATATG 1008
DB      305 TyrLeuArgTyrPheGluGluIleThrAlaGluLysPheCysArgLysArgAsnLeuLysLeu 324
QY      1009 AAAAAGTCAAGGACTCTGTCGT-----AATGCAAAAGACCTTATATTTAGT 1059
DB      325 GlnAsnAlaIleLysAsnCysHisGlyMetAspAspAspGlyLysGluLysTyrCysSer 344
QY      1060 CATATGACATGATGATTTGACACACATATTTGAAAAAGTATTTGATTTGGATTAAT 1119
DB      345 ArgAsnGlyTyrAspCysThrLysThrIleArgSerIleAspLysTyrSerMetAsnArg 364
QY      1120 AAGTACTGCTGCTGTGACATAATGCAAAAGTTTGAAGTTTGTTAGGAAATCAACAA 1179
DB      365 GlnCysThrLysCysLeuTyrValCysAspProTyrValLysThrIleAspAsnLysLys 384
QY      1180 GAACATTTTAAAAAACAAGAAAGAAATATGAAGAAATATACATATTTATGCAAC 1239
DB      385 LysGluThrGluLysGlnLysLysCysGlnLysGlnLysIleTyrArgAsnAsnGlnSer 404
QY      1240 GATACAAATTTGCAATTAATATTAATAGTAATATTAAGAACATTTTGAAGAAACCT 1299
DB      405 SerGlnAsnSerProLysAsnTyrAsnAsnMetCysGluThrAspPheTyrGlyAsnLeu 424
QY      1300 AAGGAAGCAATATATGACACTTATTAATTTTAAATTTACTAATTAAGAGAAAGTAT 1359
DB      425 LysLysAspTyrGlnSerMetAsnAsp---PheLeuLysLeuLeuAsnSerGlnThrPro 443
QY      1360 TGTAAAGAGATTAACAGAGAAAGAGATATTAATTTACTTAACAGTGTGTGACAA 1419
DB      444 CysThrAsnIleIleAspGluLysArgLysIleAspPheThrLys-----AspProGlu 461
QY      1420 GGGATATTTATGTTGACAAATATTTGCCAAGTGTGCCGACTGCGGGGTCAAA----- 1473
DB      462 GlnThrPheSerHisThrGluTyrCysAspProCysProThrPheGlyLeuLysGln 481
QY      1474 TGTGATGTTTAAATTAACACACAAATGCAATGATTAAGATGCGTGAATTAATGA 1533
DB      482 AlaAspGly---ThrTrpLysArgGluTyrGluAsnAspProGlnLysProIleLysPro 500
QY      1534 GACTATTAACCTCCAGGAGGTGAAGCTTACTAATATCACTGTCTTATAGTGGTAT 1593
DB      501 LysTyrGluThrProLysGlnGluLysGluProThrGluIleAspValLeuTyrThrGlyLys 520

```


Db 1204 AsnProProProGlnIuThrLysAspAlaCysIleCysAsnThrArgProProLysGlu 1223
QY 3736 GTACGATTAAGCAAGAAAAA-----AAAGCG 3762
Db 1224 AsperSerArgSerArgSerGluAspSerAspGluGluLysValLysVal 1243
QY 3763 GAAGAAAAGATTCGAA-----3780
Db 1244 GluGluLysAlaThrGluAspAlaValAspThrGlyProProProAlaProLysGluAla 1263
QY 3781 -----TGTAAACAGTGAATGATTAATCTTAAAGAAAAAGTGAAG 3822
Db 1264 ThrThrThrLeuAspValCysProIleValAlaGlyValLeuThr-----Lys 1279
QY 3823 AAACAGTGAAGATTCATCCAAAAAGCAATGATGATATCC---GATTCGCA 3879
Db 1280 GluAsnLeuGluAsnAlaCysProThrLysTyrGlyProLysAlaProThrSerTyrLys 1299
QY 3880 TGC-----GGAATATA 3891
Db 1300 CysIleProThrGluLysThrAsnAlaAlaThrGlySerGluGlySerSerGlyAsnGly 1319
QY 3892 AATTGATGGAAGACCCCTCT-----3912
Db 1320 AlaLeuGluArgAlaLysArgAlaThrValGluSerGlySerProValThrSerAsnSer 1339
QY 3913 -----GTGTATGCCCCCTAGACAGCAAAAGTATGCGTA-----CATTTCTTG 3957
Db 1340 GlySerIleCysIleProProArgArgArgLeuTyrIleGluLysLeuHisAspTrp 1359
QY 3958 GCAAATGATTAATCAATAATAATAATAATCAATCAATCAATGTT-----3996
Db 1360 AlaSerGlyAsnThrValValSerGlyGluAlaGlnThrProGluGlyLysThrSerSer 1379
QY 3997 -----AATTAAAGAGCTTTCATCAATCTGACACA 4029
Db 1380 ProSerGlyLysGluThrProSerAspLysLeuArgThrAlaPheIleGlnSerIleAla 1399
QY 4030 GCAGAAACATCTCTCTCATGATTTATTAATAAGTAAAGATGATGGAAGAAATCACTC 4089
Db 1400 IleGluThrPhePheLeuThrAspArgTyrLysGlu-----LysGluIle 1415
QY 4090 GATTAAGATTAATAAGAA-----4107
Db 1416 GluLysLysGluLysLysValAlaAsnGlyLysLeuValProSerLeuAsnGlyPro 1435
QY 4108 -----GGCAAA 4113
Db 1436 ProGlnGlnProGlyValThrGlyAspSerProGlnSerLysLeuGlnGlnThrGlyVal 1455
QY 4114 ATTCCTCCGATTTTGGATTCATGTTCTACACATTTGGAGATTTATGAGATTTTGA 4173
Db 1456 IleProProPhePheLeuArgGluMetPheThrLeuGlyAspTyrAlaAspIlePhe 1475
QY 4174 TTGGACACAGT-----ATATCAAAAGCTATGCTGAG-----4206
Db 1476 PheGlyLysAsnAspIleValIleAspThrLysAsnGlyLysAspIleAlaGluArg 1495
QY 4207 GGAAGTAACTAAAGCAATAGATTCCTTTCATTAATAATGTGACCAAAATCTCCT 4266
Db 1496 GluLysLysIleLysAspAlaIleGluArgValLeuLysAsnAlaAspSerGlnProPro 1515
QY 4267 AATGGAAGAACCGCAAGATGCTGACAGACATAGTCATGATGATGGAAGATGAT 4326
Db 1516 SerAspGluLysArgGlnThrTrpTrpGluGlnAsnGlyLysIleThrPheAsnGlyMet 1535
QY 4327 CTATGTCCTACTAGA-----MAAAT 4347
Db 1536 IleCysAlaLeuThrTyrLysGluLysAspGluLysGlyThrProLeuLysGlnAsnGlu 1555
QY 4348 GGGGCAAAA-----AAAGATGATTTTACCGAAACTACGG 4383
Db 1556 GlyLeuLysSerAlaLeuThrAspGluLysAsnLysLysProLysAspGlnLysTyrGln 1575

QY 4384 TACAAACAGTCAAAATTTAGTCAAAAGC-----4413
Db 1576 TyrAspLysValLysLeuAspGluAsnSerGlyThrSerProLysThrAsnAspHisVal 1595
QY 4414 -----ACCACTTTGGAGAAATTTGCCAAACAGACCCAGTTTTCAGATGGCTAACCGAA 4467
Db 1596 ProProThrProLeuThrAsnPheIleSerArgProProTyrPheArgTyrLeuGluLys 1615
QY 4468 TGTACAGCAGTATGCTATACAGCAAAAAATTTGAAGAGATGTCAGGAAAAATGT 4527
Db 1616 TrpGlyGluThrPheCysArgGluArgLysLysArgLeuGluLysIleLysValGluCys 1635
QY 4528 AAGTCAATGAC-----CAATTGAATGTGAT-----4554
Db 1636 MetAspGluAspGlyLysLysGlnLysCysSerGlyAspGlyLysAspCysGluGluIle 1655
QY 4555 -----ACGAATGTAATTAAGAAATGC 4575
Db 1656 ArgLysGlnAspTyrSerThrValArgAspPheTyrCysProGluCysGlyLysTyrCys 1675
QY 4576 GAGCACTACGTTAAATATATGAAAAA-----GAGTCGATTCACAAAGATTAATAT 4632
Db 1676 ArgPheTyrLysArgTrpIleGluLysLysAspGluTyrAspLysGlnLysGluAla 1695
QY 4633 TACAAGCATGAACGCGCAAAAAAAGATGATGACACACATGCTGTAATGTTACA 4692
Db 1696 TyrAsnAsnGlnLysThrAspAlaArgArgAsnAsnAspAsnAlaPheSerThrThr 1715
QY 4693 GACTATACGTGAACAGATGCAACAGATTAATCTGTAACAGAAATTAATGCTAGTGT 4749
Db 1716 LeuAspThrCysThrThrAlaGlyAspPheLeuGlnThrLeuLysAsnGlyProCysLys 1735
QY 4750 -----GGTGAATAGCCTGGAAGTCCCTGCTGTGATCAAAAGAAATATACAAATG 4797
Db 1736 AsnAspAsnValAspAspSerGlyLysAsnLysLysIlePheAspGluAsnGlyAspThr 1755
QY 4798 TTGAAGAAACAGGCTCTATGATGCGCAACATATGCTGGGCGCAAAATTTATGAA 4857
Db 1756 PheLysTyrThrThrGlnTyrCysGlyThr-----CysSerLeuAsnGlyPheLysCys 1772
QY 4858 AATGACACAAA-----TATACTACATTTGAGTAAAGATGACCAAGATTA 4908
Db 1773 AsnGlyAspAspCysArgValArgThrAsnValThr-----CysAsnGlySer 1788
QY 4909 GTAAAGCAGCAAAACACAGCTATTAAAGTGGCAAAACAAAGCTCTAATACATCAAT 4968
Db 1789 AsnArgThrThrThrIleThrAlaAspAspIleLysAsnGlyLys-----AsnSerAla 1806
QY 4969 AACTGGAAGAAATGACTGAAGATGCTTTTCTCTCTGCTGACTAGATATATGTTT 5028
Db 1807 GluIleAsnMetLeuValSerAsp-----1814
QY 5029 CAGCAATGAGTGGCAATTAATACAGATCCAGAAAGTTAAAGATGAAGATGGTTCGAAAA 5088
Db 1815 -----AspIleAsnSerGlyAsnGlyPheAsn---1823
QY 5089 AGATGATGGAAGTGGCGCAACGAAGGTTCAATTTGGTCAATACACAAAGAAAA 5148
Db 1824 -----AspLeuGlnLysCysLysAsnAlaAsnIlePheLysGlyLysGluAsn 1840
QY 5149 AAAGAAAAGAGAAAATAAAAGCTGCGATGCGCAAAATATTTTAAAGAGGTCCGCT 5208
Db 1841 LysTrpLys-----CysValTyrPheCysLysSerAspAla 1852
QY 5209 TGTAGTGTATGAATATATGTTTATGATTAAGATATATATCTA-----5256
Db 1853 CysGlyLeuLysLysAsnAsnAspIleAspGlnAsnGlnIleIleLeuIleArgAlaLeu 1872
QY 5257 -----GATATGATAATTTGGAAGATGAAAAAAGACCGAGAAATTTGAG 5307
Db 1873 PheLysArgTrpLeuGluTyrPheLeuAspAspTyrAsnLysIleArgLysLysLeuAsn 1892

```
QY 5308 AAAATATTACAAAAATGACATCAGTTGGCAAGAGTATAGTACTACAGCAAT 5367
Db 1893 ProCysIleAsnAsn-----|||:::|::|::|
QY 5368 CCGGTAAGTACTCGCGAAAAATTTTCTGGACGAAAAATAGAGATGTGTGGACGCA 5427
Db 1908 ValGluGln-----|||::|::|::|
QY 5428 ATGATATGGGGGTACAAACGTGTAGGATGATGAAATATAGTGCAGAGAT 5487
Db 1921 -----AsnIleLysAsn-----ArgPhe 1926
QY 5488 GATGAAGATCTAAAAAATGTGTTCTGTACTTCAGATGATCATTCCTATGGGAAA 5547
Db 1927 AsnGluGlnTyr-----:::|||||::|::|
QY 5548 AATCGGATGAAGTACTCGGTACTCAGTTCTTCGATGGTTTCCGAAATGGGTGAAGAT 5607
Db 1939 SerSerPheArgSerPheLeuValAspLeuIleArgGlnIleAla-----AlaThr 1955
QY 5608 TTTTGCAACATTAAGAAAAAGAAATTGGAAATTGTAGGGCG-----TGTAAATGAT 5661
Db 1956 IleAspLysGlyAsnHisAsnGlyLeuValLysSerValLysCysAsn--- 1974
QY 5662 TATACTTGTGGTATATGAGATTAAGAAAGAAATGTACAGATCGGTACACATAT 5721
Db 1975 -----CysGlyAsnAsnSerGlnAsnGlyLysGlu----- 1984
QY 5722 AAAAAATTATTAGTAGTGAAGCAACACAGTATGAAAAACAAATCAAAAAATATGGTGA 5781
Db 1985 -----GlyGlu 1986
QY 5782 AATTAAGACAAAAATATATCCAGCATCTGTGGCAAAAGATGACAGACGCTCGCAA 5841
Db 1987 GluAsnAspLeuValLeuLys-----LeuLeuGln 1996
QY 5842 TATTAGACAAACAAATTAATAAAATTTGTGAAT-----AAAAGTGAGATTGTGAA 5895
Db 1997 LysLeuGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2016
QY 5896 TATTAAGTATGTAAGATGTGTCCACACAGCGATTAAGTAT----- 5937
Db 2017 GlnProCys-----GluValSerProAsnHisIleGluAspValLysGlnProLeuGlu 2034
QY 5938 -----GTAATAGTCAAAATATATCCCGCATCTTAGACGAT-----GAACCA 5979
Db 2035 GluGlnGluAsnThrValGluHisProLysIleCysAspAspValLeuLysThrGlnPro 2054
QY 5980 AAAGAAGTTGAAGAAAGTGAATTTGTCAAGTCCACAGAGTCCACACGCTGTACGAGG 6039
Db 2055 GlnProGlnGluProGlnLysLysLysLysLysLysLysLysLysLysLysLys 2074
QY 6040 GAACACACGCTACACGCGGTATCATATATCAAAAGCGACGCGCATGCAAAAAGAGCG 6099
Db 2075 GluGlnGluGlnGluLys-----LysGluGlnLysAspLysGlnGlnGlu 2089
QY 6100 AAAACAGCGCGCGCTACAAACAGCGCAAAAGTGAATCTTAACACAGAAATGCGA 6159
Db 2090 ProGlyLeuProProProProProProProProProProProProProProProPro 2108
QY 6160 GCACAAACAGCAACCGACGACGACGACGACGACGACGACGACGACGACGACGAC 6208
Db 2108 oProLysProLysProProProGlnValGluLysAsnProTyrGlnHis 2124
```

Search completed: April 28, 2003, 11:59:58
Job time : 1665.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 28, 2003, 10:53:21 ; Search time 129 Seconds

(without alignments)
6834.270 Million cell updates/sec

Title: US-10-087-013-1

Perfect score: 19455
Sequence: 1 atgggggtctctctgcaata.....atatatgatagaattatg 10628

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame-n2p.model -DEV=xlp
-O=cn2.1/USPTO.spool/US10087013/runat.28042003.102845.20254/app.query.fasta.1.10823
-DB=SwissProt.40 -OPMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPTC=0
-LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human4.0.cdi
-LIST=45 -DOCALLIGN=200 -STRAT=1 -END=1 -MATRIX=biosum62 -TRANS=human4.0.cdi
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10087013.eccn.1.1.224.0/runat.28042003.102845.20254 -NCPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-NARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEX=7

Database : SwissProt.40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	618	3.2	1435	1	EBAL_PLAFC
2	435	2.2	2869	1	RBP1_PLAYB
3	420.5	2.2	6669	1	NEBU_HUMAN
4	384.5	2.0	3135	1	S230_PLAFO
5	379	1.9	3210	1	CENF_HUMAN
6	349	1.8	1153	1	PVDB_PLAKN
7	344.5	1.8	1070	1	PVDR_PLAKN
8	343	1.8	1875	1	MLPI_YEAST
9	337.5	1.7	1070	1	PVDB_PLAKN
10	337.5	1.7	1073	1	PVDB_PLAKN
11	329	1.7	2116	1	MYS2_DICDI
12	322.5	1.7	2748	1	NUM1_YEAST
13	319.5	1.6	1391	1	MST2_DROH
14	313	1.6	2663	1	CENE_HUMAN
15	311	1.6	2492	1	ATRX_HUMAN
16	311	1.6	2704	1	ATRX_HUMAN
17	302	1.6	2476	1	ATRX_MOUSE
18	299.5	1.5	2339	1	RPCI_PLAFA

19	298.5	1.5	1956	1	ATX1_PLAFA	004956	plasmodium
20	290	1.5	1957	1	YD86_SCHPO	Q10411	schizosacch
21	289	1.5	1630	1	MSPI_PLAFC	P04932	plasmodium
22	289	1.5	1639	1	MSPI_PLAFC	P04932	plasmodium
23	289	1.5	1639	1	MSPI_PLAFC	P04932	plasmodium
24	287	1.5	3418	1	DMC_CANPA	097592	canis fam1
25	287	1.5	3418	1	DMC_HUMAN	P51587	homo sapien
26	279	1.4	3911	1	AKA9_HUMAN	P11532	homo sapien
27	274	1.4	2230	1	GOG4_HUMAN	Q99936	homo sapien
28	271.5	1.4	1790	1	USO1_YEAST	Q13439	homo sapien
29	270.5	1.4	1805	1	HMW2_MYGE	P25386	saccharomyc
30	270.5	1.4	3678	1	DMD_MOUSE	P47460	mycoplasma
31	269.5	1.4	2022	1	ANT1_ONCVO	P11531	mus musculu
32	268	1.4	1658	1	YME7_YEAST	P21249	onchocerca
33	267.5	1.4	5430	1	ACF7_HUMAN	Q03661	saccharomyc
34	265	1.4	1679	1	YIO9_YEAST	Q9UP63	homo sapien
35	263.5	1.4	1726	1	MSPI_PLAFC	P04932	plasmodium
36	262.5	1.3	5337	1	ACF7_MOUSE	Q9GX80	mus musculu
37	261.5	1.3	1726	1	MSPI_PLAFC	P50435	plasmodium
38	259.5	1.3	3660	1	DMD_CHICK	P11533	gallus gall
39	259	1.3	1557	1	DVAL_DICVI	Q24702	dictyosacch
40	259	1.3	1727	1	ALM1_SCHPO	Q9UKF5	schizosacch
41	258	1.3	2871	1	DESP_HUMAN	P15974	homo sapien
42	255	1.3	1939	1	MYH6_MESAU	P13529	mesocricetu
43	252.5	1.3	1928	1	MYH1_YEAST	P08964	saccharomyc
44	245	1.3	2198	1	YIJ2_CAEEL	P34367	caenorhabdi
45	244.5	1.3	1701	1	MSPI_PLAFC	P13819	plasmodium

ALIGNMENTS

RESULT 1	ID	EBAL_PLAFC	STANDARD	PRT	1435 AA.
AC	P19214				
DF	01-NOV-1990 (Rel. 16, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	01-FEB-1996 (Rel. 33, Last annotation update)				
DE	Erythrocyte-binding antigen EBA-175.				
OS	Plasmodium falciparum (isolate Camp / Malaysia).				
OC	Eukaryotes; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxId=5835;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=90377299; PubMed=2204835;				
RA	Sim B.K.L.;				
RT	"sequence conservation of a functional domain of erythrocyte binding				
RL	antigen 175 in Plasmodium falciparum."				
CC	Mol. Biochem. Parasitol. 41:293-296(1990).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
DR	EMBL; X52524; CA36756.1; -				
KW	Antigen.				
FT	DOMAIN 159 1104				
FT	-----				
FT	VARIANT 1031 1031				
FT	SEQUENCE 1435 AA; 167389 MM; 3244309021BIC3D6 CRC64;				
SO	-----				
SO	ESSENTIAL FOR BINDING TO				
SO	ERYTHROCYTES.				
SO	E -> V (IN STRAINS FCR-3 AND ITG).				
Alignment Scores:					
Pred. No.:	5.4e-22				
Score:	618.00				
Percent Similarity:	36.80%				
Best Local Similarity:	22.20%				
Query Match:	3.18%				
DB:	1				
	Gaps:				
	72				

```

US-10-087-013-1 (1-10628) x EBA1_PLAFC (1-1435)
OY 3301 GAAGCTTGAATGATATATATAGATATATAGCAAGAAACA-----TCG 3351
Db 20 LysAlaArgAsnGluTyr--AspIle-----LysGluAsnGluLysPheLeu 34
OY 3352 AATATAATATCATGATTAATACAAAGATTTACATGACAAAGCAAAAGCTGTAGTAAAT 3411
Db 35 AspValTyrLysGluLysPheAsnGluLeuAspLysLysTyrGluValGluLys 54
OY 3412 AGTGGTATTTAGAGCTTCAGTACTGCCAAAATATATAGCAGAGATGTTATGAATTT 3471
Db 55 Thr-----AspLysLysIlePheThrPhe 62
OY 3472 TTG---TCGGAATTTATACCAAAATAGTGGCAAAAGTATTAAGT----- 3516
Db 63 IleGluAsnLysLeuAspIleLeuAsnAsnSerLysPheAsnLysArgTyrLysSerTyr 82
OY 3517 GGTACTAGTGTATGAAAGCGCTGATTCGTTACTACACCGATATGAAATGTTGGAGCA 3576
Db 83 GlyThrProAspAsn-----IleAspLysAsnMetSerLeuIleAsn----- 96
OY 3577 TATCTCATGATACGAAATTTTGATGATTTGTCAGTACAAATGAGTTTGTGATGA 3636
Db 97 ---LysHisAsn--AsnGlu---GluMetPheAsnAsnAsnTyrGlnSerPheLeuSerTh 114
OY 3637 AAAAGTATGTTAGAGCAATTAACGAAATATGCTTTAGAGATTAACCCAGGACCATGAT 3696
Db 114 rSerSerLeuIleLysGlnAsnLysTyrValProIle-----AsnAlaValArgVal 131
OY 3697 GGTGGCTGTGGTTGTAAAGTTCGAAACGCAAGAGGTACAGATTAAGCAAAAAA 3756
Db 131 lSerArgIleLeuSerPheLeuAsp----- 139
OY 3757 AAAGCGAAGAAAAGATTCGGAATGTAACAGTG-----ATGATATACTTAA 3806
Db 140 -----SerArgIleAsnAsnGluArgAsnThrSerSerAsnAsnGluValLeuSe 156
OY 3807 AGAA---AAGCATGGAAGAAAGAAACAGTAGAA---GATTCATCCAAAAAGAAATGTAA 3860
Db 156 rAsnCysArgGluLysArgLysGluMetLysTrpAspCysLysLysAsnAspArgSe 176
OY 3861 TGGATATCCCGATTTGGCAATGCGGAATATTAATTTAGGAAGACCTCGTGTGTAT 3920
Db 176 rAsnTyr-----ValCysTil 181
OY 3921 GCCCCTAGAGACAAAGTATGCTATCTTTCGCAAAATGATTAAGAAATAAAAA 3980
Db 181 eProAspArgArgIleGlnLeuLysIleValAsnLeuSer-----IleIleLeu 197
OY 3981 ATTACATCAACAAGTTAATTTAAAGACCTTTTCATCAATCTGCAGCAGCAAGAAACAT 4040
Db 197 sThrTyrThrLysGluThrMetLysAsnAspPheIleGluAlaSerLysLysGlu- 215
OY 4041 CTTTCATGCTGTTTATTAATAAGTAGAGTGTGAAGGAATGAATCGATTAAGAATT 4100
Db 216 ---SerGlnLeuLeuLysLysAsnAspAsnLysTyrAsnSer----- 229
OY 4101 AAAAGAGGCAAAATTCCTCCCGCATTTTTCGATCCCATGTTCTACACATTTGGAGATTA 4160
Db 230 -----LysPheCysAsnAspLeuLysAsnSerPheLeuAspTyr 242
OY 4161 TAGAGATTTTATTTGGAACAGATATATCAAAAGTCATGCTGAGGAGCAAGAAACATAA 4220
Db 242 rGluHisLeuAlaMetGluLysAsnAspMetAspPheLys---GlyTyrSerThrLysAlaGlu 261
OY 4221 AGAGCAATATAGTTCTCTTTCAAAATAGTGACCAAAATCTCCTAATGGAAAAACA- 4278
Db 261 uAsnLysIleGlnGluValPheLysGluAlaHisGluLysIleSerGlnHisLysIleLys 281
OY 4279 -----CGCCAAAGATGTGAGCAAGAACATAGCATGAGATATGGAGATGATG 4331
Db 4331 ----- 4331

```

```

Db 281 sAsnPheArgLysGluThrTrpAsnGluPheArgGluLysLeuTrpGluAlaMetLeu- 300
OY 4332 TGCATAGTAAAAATGGGCAAAAAAGATGATTTTACGAAAGTACAGGTACAAACA 4391
Db 301 -----SerGlnHisLysAsnAs 306
OY 4392 CGTCAAAATTATAGTACAAAAAGCAACCACTTTGGAGAAATTTGCCAAAGCAAGTTT 4451
Db 306 nIleAsn---AsnCysLysAsnIleProGlnGluLeu-----GlnLeuTh 321
OY 4452 ACGATGGCTAACCGAATGAGTACAGCAGCTATTTCATACAGCAAAATATTTAGAGA 4511
Db 321 rGlnTrpIleLysGluThrPheIleGluThrLeuLeuGluLysAsnArgSerLysLe 341
OY 4512 TGTGCAGGAAAAATGTAAGTCAAAATGCAATGAAAG---TGTGATACAGATTAATTA 4568
Db 341 uProLysSerLysCysLysAsnAsnThrLeuTyrGluAlaCysGluLysGluCysIleAs 361
OY 4569 GAAATCGAGGACTACGTTAATAT--ATGAAAAAAGAGTGGATTCACAAAGA 4625
Db 361 ProCysMetLysTyrArgAspTrpIleLeuArgSerLysPheGluThrPheIleuSe 381
OY 4626 TAAATATTCAAGATGAAAGCGACAAAAAAGATTCGATAGACAAACATTCGTGAT 4685
Db 381 rLysGluTyrGluThrGlnLysValProLysGlu----- 392
OY 4686 GGTTCACAGCTTACTGGAACAGATGCAACAGATTTAGTGAACGAAATTTACTGTAG 4745
Db 393 -----AsnAlaGluAsnTyrLeu---IleLysIleSerGluAs 404
OY 4746 TTGTGTGATTAAGCCTGGAAGTCCCTGTGGTACAAAGAAATATCAATTTAGAAA 4805
Db 404 nLysAsnAsp-----AlaLysValSerLeuLeuAs 415
OY 4806 ACAGCTTACTATGATGCCGCAAAACANTGCGTGC-----ACAAAATTAATGA 4856
Db 415 nAsnCysAspAlaGluTyrSerLysTyrCysAspCysLysHisThrThrIleValLys 435
OY 4857 A-----AATGAC-----GACAAATATCTAATCAATTTGAGATA 4889
Db 435 sSerValLeuAsnGluLysAsnAspAsnThrIleLysGluLysArgGluHisIleAspLeuAs 455
OY 4890 AGATAGTGCAAA---GGATAGTAAAGAGGCA---AACACAGTCTATTAAAGTGCA 4943
Db 455 rAspPheSerLysPheGluCysAspLysAsnSerValAspThrAsnThrLysValTrpGlu 475
OY 4944 AAACAAAGGCTCTAATACATCAATTAATCTGAAAGAAATGCAAGACAGTCTTTCC 5003
Db 475 uCysLysAsnProTyrIleLeuSer-----ThrLysAspValCysValPr 490
OY 5004 TTCTCGTACGACTAGCTATATGTTTTCATGCAATTCGATGGCAATTAATACAGATCCAGAGT 5063
Db 490 oProArgArgGlnGluLeuLysCysLeuGluLysAsnIleAspArgIleTyr----- 505
OY 5064 TAAAGATGAAAATGGGTTG-----CGAAAAGATTTGGAAGTGGCGCAAGGAGG 5117
Db 506 ---AspLysAsnLeuLeuMetIleLysGlnHisIleLeuAlaIleAlaIleTyrGluSe 524
OY 5118 GTACAAATTTGGGCAATACATACAAAGAAAGAAAAAGAAAAAGAAATTAACAAACGTCGGA 5177
Db 524 rArgIleLeuLysArgLysTyrLysAsnLysAspAspLysGlu----- 538
OY 5178 TCGGCACAAATATTCTTANGAGGTCCCGCTTGTAGCTATGAATATATGTTTATGA 5237
Db 539 -----ValCysLysIleIleAsnLysThrPheAlaAs 549
OY 5238 TTTAAGATATATTTCTAGTATGATTAATTTGGAAGTGAAGAAACAAAGACGAGGA 5297
Db 549 rPheArgAspIleIleGluLysThrAspTyrTrpAsnAsp-----LeuSerAsnArg 566
OY 5298 AAATTTGAAGAAAATATTTAAACAAAAATGGAACATGATGCGAAAGAAAGATATGATG 5357
Db 566 gLysLeuValGluLysIleAsnThrAsnSerLysTyrValHisArgAsnLysLysAsn- 585

```

```
QY 5358 TACAGAAATCCGGTAGTACTCGCGGAAATTTTCTTGAGAACGAAATATGAAATGTGT 5417
Db 586 -----AspLysLeuPheArgAspGluTrpTrpLysValIleLysLysAspVa 601
QY 5418 GTGGAACGCAATATATCGCGGTACAAACGTGGTAGGATGATGAAATAGTGGAAATAG 5477
Db 601 LrrpAsnValIleSerTrpValPheLys----- 610
QY 5478 TGCAGAGATGATGAGATCTAAATAAATGTGTTCTGTACTCTTCAGATGATGATATTC 5537
Db 611 -----AspLysThr 613
QY 5538 TATGGGAAATATCGGATGAGTACTCGGTATCAGTTTCTTCGATGTTTCCGATG 5597
Db 613 rValcLysGluAspAspIleGluAsnIleProGlnPheArgTrpPheSerGluTr 633
QY 5598 GGGTGAAGATTTTGCAGCAATATAGAAAGAAAGATTTGGTAGGGCGGTAA 5657
Db 633 pGlyAspAspTrpLysGlnAspLysThrLysMetIleGluThrLeuLysValGluCysL 653
QY 5658 TGATTTACTTGTGTATATGATATGATAAAGAAAGATGTACAGATCGGTACACA 5717
Db 653 sGluLysProCysGluAspAsp-----AsnCysLysSerLysCysAsn 668
QY 5718 ATTTAAAAATTTATATGATGATGGAACCAAGTATGAAACAAATCAAAAAATATGG 5777
Db 668 rTyLysGluTrpLysSerLysLysGluLysGluLysAsnLysGlnIleLysGlnTyG 688
QY 5778 TGGG-----AATTAAGCAAAATATATTCGACATCCCTGTGCAAAAGATGC 5825
Db 688 ngLutTyGlnLysGlnAsnTrpLysMetTyLysSerGluPheLysSerLleLysProG 708
QY 5826 AGAGAGCGCTCGGCAATATTTAGCAAAACAATTAATAAATTTGTGAAATTAAGTGG 5885
Db 708 uValTyLysLeuLysLysTySerGluLys-----CysSerAsnLeuAsnPh 723
QY 5886 AGATTGTGAATATAG-----TGT-- 5904
Db 723 eGluAspGlnPheLysGlnLysLeuHisSerAspTyLysAsnLysCysThrMetCysPr 743
QY 5905 -----ATGAAAGATGTGTCCACACAGCATTAAGTATAGTAAATATATGCCCCG 5960
Db 743 oGluValLysAspValProLysSer-----LleIleArgAsnAsnGluIleThrSerGln 762
QY 5961 ATCATTTAGACATGATCAACCAAAAGAAATTT-----GAAGCAAGTGTATTTGCA 6008
Db 762 uAlaValProGlnGluAsnThrGluIleAlaHisArgThrGluThrProSerLleSerG 782
QY 6009 AGTCCACGAGGTTCACACAGTGTAGCAAGGAAACACCGTCACACAGGATACATGAT 6068
Db 782 uGlyProLysGlnLysGlnGluLysGluArgAspAspSer-----Le 797
QY 6069 ATCAAAAAGCGAGCGCAAAAAGAAAGCAAGCGCGCTCAAAAACAGCCGGA 6128
Db 797 uSerLysLysSerValSer-----ProGluAsnSerAsnPro-- 809
QY 6129 AAAAGTGAATCTTACCAACAGAAATGCGAGCAAAACAGAACCCGACAGCAGACA 6188
Db 810 -----G 810
QY 6189 ACAAAACGAAAGCAATCAACAGCAACAGCAACAGAAATTCGACTGGCC--ACAAT 6245
Db 810 urHisAspAlaLysAspThrSerAsnLeuLysLysLeuLysGlyAspValAspLysSerMe 830
QY 6246 GGTAAAGCCATCTTTCGATTAACACAGATAGCAGGGGGAATTAAGGTTGTATCC 6305
Db 830 tProLysAlaValIleLysSerProAsn-----AspAsnIleAsn-- 844
QY 6306 AAAAAGCTATGACAAATATCTAAATGGGTTGATTATGTAGTAAGTCTAAAGAAATGA 6365
Db 845 -----ValThrGluGlnGluLysAspAsnI 852
QY 6366 AAATGCAATATGATGCTCTCTAGGAGAAAAAATTTATGTAATATATATACAAATATT 6425
Db 852 eSerGlyVal-----AsnSerLysProLeuSerAspAspValArgProAs 867
QY 6426 AAATTTATGAAACGTGAATAAGCGGTGACATGATATATAAGAGCGCTTTTATTAATGTC 6485
Db 867 pLysLysGlnLeuGlnAspGlnAsnSerAspGluSerGluGlnThrValVal----- 884
QY 6486 AGCAATGAAACCTCAATTTTGTGTTAAATATATATATGAAATATCTGCACAGAAAA 6545
Db 885 -----AsnHisLleSerLysSerProSerIleAsnAs 895
QY 6546 TGAATTCGAAATGCAACAAATTCAGATGAAATTAAGAATATGATTAACATATGG 6605
Db 895 nGlyAspAspSerGlySer-----GlySerAlaThrValSe 907
QY 6606 TGATTTAAGATATGTTTTTTGGAAGTATTTCTATGATTAATAAATATATATGCT 6665
Db 907 rGluSerSerSerSerAsnThrGluLeuSerLleAspAspAspArgAsnGluAspThrPh 927
QY 6666 AACCAATAGTGTAAACCAACATTCATCAATGAATAATATAGAAAAAACGAGTAAAAAAA 6725
Db 927 eValArgThrGlnAspThrAlaAsnThrGluAspValIleArgLysGlnAsnAlaAspLy 947
QY 6726 AGATGAAGAAATTCGTAATAATTTTGGGAGAAAAATTAATAATTTTATTTGGCAGCAAT 6785
Db 947 sAspGluAspGlnLysGlnLysAlaAspGluLysArg----- 958
QY 6786 GATPATGATTAACATATCATCATCTCACAGACGAAAC-----GAAAAAGAAA 6833
Db 959 -----HisSerThrSerGluSerLeuSerSerProGluGluLysSme 972
QY 6834 AATTAGAGATATTTACCGATACATGACATGACCAACATGACCGCTTCCTTGAGAGATT 6893
Db 972 tLeuThrAspAsnGluGlnLysAsnSerLeuAsn-----HisGluG 986
QY 6894 TGTAAAAAGCCCCAATTTTGGATGTTGCACGAATGGCCACAGCAATTTGTAATTA 6953
Db 986 uValLysGlnHisThrSerAsnSerAspAsnValGlnGlnSerGlyLysIleValAsnMe 1006
QY 6954 GAGCAAGCAACGTTGTTAAA-----TTGAGCGCGGG 6986
Db 1006 tAsnValGlnLysGlnLysLeuLysAspThrLeuGluAsnProSerSerLeuAspGluG 1026
QY 6987 CTGTAAAGCAATATGATGATGATGATGAT----- 7017
Db 1026 yLysAlaHisGlnLysLeuSerGluProAsnLeuSerSerAspGlnAspMetSerAsnTh 1046
QY 7018 -----GACGCTAAGACACAAAGATGTGCAGAGCGCTGTATACATATCAAAA 7064
Db 1046 rProGlyProLeuAspAsnThrSerGlnGluIleThrGluArg-----IleSerAsnAs 1064
QY 7065 TTTTATTAAGAGTGAAGAACTCAATATGAAGCAAAAGAAAGTTCAAAAAGAT-- 7122
Db 1064 ngLutTyLysValAsnGlnLysArgGluAspGluThrLeuThrLysGluTyGluAspTr 1084
QY 7123 -----AAAGATGCAAAAAATGATATATCC 7151
Db 1084 eValLeuLysSerHisMetAsnArgLysSerAspGlyGluLeuTyArgGlu----- 1102
QY 7152 TTTACTGAAAGACATAGAGAAAGCAACATGTGCTCATGATATTTAAACATGAAT 7211
Db 1103 -----AsnSerAspLeuSerThrValAsnAspGluSerGlnAspAlaGluAlaLysMe 1120
QY 7212 AAAGAAATATATGTCGATATAGATTTCTTGTATGCAAAAACCTTCTTCAACAATAC 7271
Db 1120 tLys-----GlyAsn--AspThrSerGluMetSerHisAsnSerSerGlnHisI 1136
QY 7272 AAAAACAACACAAATCACAATATCATGATGATGATATATGTCAGAAATGCTGAT-- 7329
Db 1136 eGluSerAspGlnGlnLys-----AsnAspMetLysThrValGlyAspLe 1151
QY 7330 -----TATGTCTCTGAAGATTTAACAGTGTGAGTGTCTCTGAACCTTCAAAA-- 7377
```


Db 1151 uGlyThrThrHisValGlnAsnGluIleSerValProValThrGlyGluIleAspGlu 1171
 QY 7378 -----AAGGATCTATGATCATACAAAAAATTAAGCTAAATAC 7424
 Db 1171 sLeuArgIleSerLysGluSerLysIleHis-----LysAlaGluGluIleAspGlu 1189
 QY 7425 TATGATATGCTAGAGAAAGCAATATTAATCTAAAGACAGAAATTAATATGGA 7484
 Db 1189 rHisThrAspIleHisLys-----IleAsnProGluLysPargAsnSerAsn 1205
 QY 7485 TATTACCTTGAGGAAAAATTTATACCTATGATCTACAAAGAAAGAAAGTAATAA 7544
 Db 1205 rLeuHisLeuLysAsp-----IleArgAsnGluIleAsnGlu 1218
 QY 7545 TAGTTGACATATATATAT 7563
 Db 1218 gHisLeuThrAsnGlnAsn 1224

RESULT 2

PL_PLAVB STANDARD; PRT; 2869 AA.
 Q00758;

DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DE 01-OCT-1996 (Rel. 34, Last annotation update)
 GN Reticulocyte binding protein-1 precursor.
 OS Plasmodium vivax (strain Belen).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 ON NCBI_TaxID:31273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE:92315338; PubMed-1617731;
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 merozoites.";
 RL Cell 69:1213-1226(1992).
 CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
 HUMAN RETICULOCYTE CELLS.
 CC -1- SUBUNIT: HOMODIMER (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Membrane-bound.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

DR EMBL; M88097; AAA29743.1;
 KW Malaria; Receptor; Signal; Transmembrane.
 FT SIGNAL 1 17
 FT CHAIN 18 2869
 FT DOMAIN 18 2807
 FT TRANSMEM 2808 2826
 FT DOMAIN 2827 2869
 FT SITE 1030 1032
 FT SITE 2599 2601
 FT SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CRC64;
 Alignment Scores:
 Pred. No.: 1.78e-13
 Score: 435.00
 Percent Similarity: 33.38%
 Best Local Similarity: 17.96%
 Query Match: 2.24%
 DB: 1
 Gaps: 129

US-10-087-013-1 (1-10628) x RBPL_PLAVB (1-2869)

QY 40 AATGACATCATATATGAGGAGATGTAAAGCCCTATATATAAGAAAGTCACAA 99

Db 428 AspAlaIleLysLysLeuGluSer-----IleAsnGluIleAspGlu 441
 QY 100 AGTCAAGAAATGTTTGGAACTTATGCCAAATATAAGCAT---CCATCAAAATAT 156
 Db 442 SerAlaGluIleAspLeuGluIleAspSerLysAlaLeuAlaAsnAsnTyr 461
 QY 157 GCA---AAAGACATGTGATTCGTTGAAAGGGATTGACGAAAGCAAGAAATTCGTGT 213
 Db 462 CysIlePheGluIleLysThrLeuAsnGluProIleLysLysAlaTyr----- 478
 QY 214 GGTCTTACGCCCATTAATTAAGCATATTAATTAATCAATTCATATTAATAGT 273
 Db 479 -----GluSerLysValIleLysSerAsnGluLeuLeuSerThrIleIleAspThr 496
 QY 274 CATAGCAACATCTAATTTACGTATGATGATGATGATGATGATGATGATGATGATG 333
 Db 497 GlyLysSerAlaThrAlaLeu----- 503
 QY 334 AGAAGAAACAAACCATTTGATGATGATGATGATGATGATGATGATGATGATGAT 393
 Db 504 -----GluIleSerThrPheAspGlu-----GluCys---AsnLysIleLysThr 518
 QY 394 -----TATTAAGAAAGAAATGATCTATAGCCGTGGCCACCTAGAAAGCATATG 447
 Db 519 GluAlaGluLysValLysAspAspAla-----GluAspIle 530
 QY 448 TGTGATTAACAC-----TTGAAGCTTAATGATTAATTAATTAATTAATTAAT 486
 Db 531 CysGluLysAsnGluGluIleLysTyrGluIleProGluSerGluAspGluThrIleAsp 550
 QY 487 CAATAATTCATGATTTATTTGGAAATGATCTACTAGTACAGCAAAATACGAAGTGAATCA 546
 Db 551 AspLysIleAsnAspLeuGluIleAspLeuIleAspGluMetLysGluTyrLysAspGlu 569
 QY 547 ATGTGTAATTAATCATTCACATTAAGAACTTCAGACGCTGTACTGCTTCGACGAAGT 606
 Db 570 IleValAsnAsnSerGluPhe-----IleSerAsnArg 580
 QY 607 TTTCGATATGAGGATATGTAAGAGATATATATGTTTAAACCAATTCATGATGAC 666
 Db 581 TyrLysAsnIleTyrGlu-----AsnLeuLysGlu 590
 QY 667 AAGTAGAAGACGGCTGTCGAGAGGTTTCAGAAATATACATGATGATGATGATGATGAA 726
 Db 591 ThrTyrGluThrGluLeuAsnAspIleGlyLys----- 601
 QY 727 GTAAAAATGATTTCAATCCTGATGATGATGATGATGATGATGATGATGATGATG 786
 Db 602 LeuGluLysAsp-----ThrSerLysValAsnPheTyrLeuMetGluIleArgLys 618
 QY 787 TGAATGTGAATGAATTAAGATGATGAGAGCATATA---ACATGATGATGATGATGAT 843
 Db 619 IleAsnThrGluLysThrLysIleAspGluSerLeuGluThrValGluLysPheTyrLys 638
 QY 844 TCT----- 846
 Db 639 GluIleLeuAspSerLysGluLysIleTyrGluLeuLysIleGluPheGluLysSerVal 658
 QY 847 -----GGATATTTATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 897
 Db 659 ThrGluIleAsnArgLeuGluAspGlyGluSerAlaArgAspLeu----- 673
 QY 898 TGGCGCCATTAACAA-----GGAAGGTTCCCTACCAATTTATGATATAT 939
 Db 674 -----HisGluGluGluIleLysGluIleLeuAspLysMetAlaLysLysValHisTyr 691
 QY 940 GTCCTCAATATTTACGT-----TGGTTCGACGATGGGAGAGA 978
 Db 692 LeuLysGluLeuLeuSerLeuLysSerValTyrPheThrGluMetAsnGlu 711
 QY 979 -----GAGTTTGGCGAAGAAATATTAATTAATTAATTAATTAATTAATTAAT 1023

Db 712 LeuLeuAsnThrAlaSerThrAspAsnMetGluGlyPheSerAlaLysLysGluLys --- 730
 QY 1024 TCCTGCTGAATGACAAAGACGCTTATATTGATCAATAATGACATGATTTGACACA 1083
 Db 731 ---AlaAspAsnAspIleAsnAlaLeuTyrAsnSerValTyrArgLysAspIleAsnAla 749
 QY 1084 ACTATTGG-----AAAAAGCTATTGCTTTGCTTTG--- 1113
 Db 750 LeuIleGluGluValGluLysPheValThrGluAsnLysGluSerThrLeuGluMetLeu 769
 QY 1113 ----- 1113
 Db 770 LysAsnGluGluMetGluLysLeuGluAsnAlaLysGluThrPheAlaLysLeuAsn 789
 QY 1114 -----GATTAATAGTACTGACTGCTGCATAATGCAAACTT----- 1152
 Db 790 PheValSerAspLysPheLysLeuThrAspValTyrThrLysMetSerAlaGluValThrAsn 809
 QY 1152 ----- 1152
 Db 810 AlaGluGlyIleLysLysGluIleAlaGluLysGluPheGluAsnValHisLysLysMet 829
 QY 1153 -----TTTGAAGTTTGGTTAGGGAATCAACAA--- 1179
 Db 830 LysGluPheSerAspAlaPheSerThrLysPheGluAlaLeuGluAsnSerMetGluGlu 849
 QY 1180 -----GAGCATTTTAAACAAACAAACAAACAAATATGAAACAAATACAA 1224
 Db 850 TyrAsnGluGluGluLysAspAlaIleGluLysHisLysGluAsnArgSerGluLysGlu 869
 QY 1225 TCCTATTATTCGACCATTAACAATTTGTCATTAATATTAAATAGTGAATTTTAAACAA 1284
 Db 870 GluTyrPheLysAsnGluSerValGluLysAspLeuSerArgLysGluThrGluGluGlu 889
 QY 1285 TTTTATGAAAACTTAAAG----- 1302
 Db 890 GluTyrThrLysHisLysAsnAsnPheSerArgArgLysGluLysLeuSerAlaGluIle 909
 QY 1303 -----GAAAGCAAA-----TATGCAACTAAT 1323
 Db 910 ThrAsnMetArgGluValIleAsnLysIleGluSerGluLeuAsnTyrTyrGluValIle 929
 QY 1324 GACACTTTTAAATTACTA-----AATGAGGAAGATATGTAAGAGAGATTA 1374
 Db 930 GluLysTyrPheSerLeuIleGlyAspGluAsnGluValSerThrAlaLys----- 946
 QY 1375 CCAGAGAAAGAGATATTACTTTTACTAACAGCTGATGACAAAGAGATATTATTCGT 1434
 Db 947 ---AlaLeuLysGluLysIleValSerAspSerLeuArgAspLysIleAspGluTyrGlu 965
 QY 1435 TCAGAAATAT-----TGCCAACTGTCTCCGACTGC 1464
 Db 966 ThrGluPheLeuGluLysThrSerAlaValGluAsnThrValSerThrIleGluInsLeu 985
 QY 1465 GGGGCTCAATGTCATGCTAATAATACACACAAATCGATATATGAT----- 1512
 Db 986 SerLysAlaIleAspSerLeuLysArgLeuAsnGlySerIleAsnAsnLysLysLysTyr 1005
 QY 1513 -----CGTGAACGTGTAATAAT 1530
 Db 1006 AsnThrAspIleAspLeuAsnArgSerLysIleLysThrLeuArgGluGluValGluLys 1025
 QY 1531 GAAGACTATAAA-----CCTCATGGGGTGTGAGACCTACTAATATCTACTGCTTTAT 1584
 Db 1026 GluMetProLysArgLysAspLysCysGlyGluAsnThrThrAlaLeuLeuLysSer 1045
 QY 1585 AGTGTGAATGACAAAGCTGATATTACACAAATAATTTTGTAAACAGCTCACT 1644
 Db 1046 LeuArgAspLysMetGlyLysIleAsnGluLysLeuAsnAspLysArgLeuAsnSerLeu 1065
 QY 1645 AATTACAAGATATAATATCAAAATGGAATGCTATTAAT----- 1686
 Db 1066 AspThrLysLysLysLeuLysPheLys-----PheTyrSerGluSerLysSerLys 1082

QY 1687 -----AAGATGATAAAT-----ATAAATAGTGAATCACTGA 1719
 Db 1083 IleHisLeuSerLysAspGluLysGlyProGluAspProLeuAsnArg-----IleAsp 1100
 QY 1720 CAATAATACGAATCAATATGATATCCCTACATA-----ATATCATTTTCAATAATTTT 1776
 Db 1101 GluTyrGluAspIleLysArgAspValAspGluLeuAsnValAsnTyrGluValIleSer 1120
 QY 1777 GAATTATGGCTTACATATTATTAAGGACTATTAAGGAAATGCAAACTTAAC 1836
 Db 1121 GluAsnLysValThrLeuPheLysAsnAsnSerValThrTyrIleGluAlaMetHisSer 1140
 QY 1837 TGTATTAATTAATCAACACGACTGATTTGATGAATGTAACGAATTCCTATGCTTT 1896
 Db 1141 HisIle---AsnThrValAlaHisGlyIleThrSer---AsnLysAsnGluIleLeu--- 1157
 QY 1897 GACGATGGCTTAACAAAGAAAGAAAGAAATGCAATATGTAAGAAACCTTCAACAAA 1956
 Db 1158 ---LysSerValLysGluValGluAspLysLeuAsnLeuValGlu----- 1171
 QY 1957 AAAAAGATATACGCAATCGTATTATAGTAATTAATAAT----- 1998
 Db 1172 -----GluAsnGluAspTyrLysLysValLysAsnProGluAsnGluLysGlu 1187
 QY 1999 -----CTTTTGAAGTTATTTTAAAGTTATGATGAATGACAAAGAT----- 2046
 Db 1188 LeuGluAlaIleArgLysSerMetSerLysLeuLysGluValIleAsnLysHisValSer 1207
 QY 2047 -----GAGCAAAATGGAAGAAACCTTAATGCAAAATATTAAGAAACAA 2091
 Db 1208 GluMetThrGluLeuGluSerThrAlaAsnThrLysSerAsnAlaLysGlyGlu 1227
 QY 2092 AATGAGTTTCCAAATTTGCAAAATATAGGACTATTAAGAAATGCAAGAACTCTG 2151
 Db 1228 AsnGluHisAspLeuGluGluLeuAsnLys-----ThrLysGlyGluMetArgAspIle 1245
 QY 2152 TTGATATCCTTAAAGAAACCTGCCAGATATGTAAGACATTAATCAACAGAGATGT 2211
 Db 1246 TyrGluLysLeuLysLysIleAlaGluGluLeuLysGluThrValAsnGluLeuLys 1265
 QY 2212 GAAACATCCCATATATGACAAACAAACCCCTGTGTAACCTGCTGAGACACCAACC 2271
 Db 1266 AspAlaAsnGluLysAlaAsnLys-----ValGluPro-----GluPro 1278
 QY 2272 ACTTAAATATTAAGAAACATATGCAATATTAAGAGCTGATACAGAGAGACGA 2331
 Db 1279 GluArgAsnIle-----IleGlyHisValLeuGluArgIleThrValGluLysAspLys 1296
 QY 2332 AAT-----CGTGGCTTTCATTAATTAAGAAAGGACACAGAGGTATATAT 2379
 Db 1297 AlaGlyLysValValGluGluMetLeuSerLeuLysThrLysIleGluLysLeuIleGlu 1316
 QY 2380 AAACGTGGGGGTGAGCAAAAGCACTTCAAGACAAATTAATGTAATATGATTA 2436
 Db 1317 GluThrSer-----AspAspSerGluAsnGluLeuValThrThrSerIleThrLys 1333
 QY 2437 CATTTAATGTAATCTGTGTTTTCATTAATGACATGATGCAAGACACAGGTGAT 2496
 Db 1334 HisLeuGluAsnAlaLysGlyTyr----- 1341
 QY 2497 GGTATACAAACAAAGATTTGCTGTAAGAACTGAATGGGAAGTGCAGAACATGCGT 2556
 Db 1342 -----GluAspValIle 1345
 QY 2557 AAAGATCAGCAAGATGTTATTAATGCTCCTAGAAAGACGACATTAATGATCAATTTG 2616
 Db 1346 LysArgAsnGluLysAspSerIleGluLeuArgGluLysAlaLysSerLeuGluThrLeu 1365
 QY 2617 GAACATTTACAAACGAGTATCAACCACTTAATGTAATGTAATGTTGATGATTAGTTAAT 2676
 Db 1366 AspGluMetLysLysLeuValGluGluValAlaAsnMetAsnLeuInsAlaIleGluGly 1385

OY 2677 AATTCCTTTTGGGGAGTCTCTATCAGCAAAATATGAAACAAGATATATAGCA 2736
 Db 1386 AsnAlaGlyIleSer----- 1390
 OY 2737 ATGTATAAGAAAGATACCTTAAGGCCCCAAAGATPACTGACCCAAACACGAG 2796
 Db 1391 -----LysGluLeuAsnGluLeuLysGlyAlaIleGluLeuLeuIleSerThrAsnTyr 1408
 OY 2797 ACAACTATCTGTGACGATACGTTACGTTTTCGACATATAGGTGATATATATTCGAGA 2856
 Db 1409 SerSerIle-----LeuGluTyrValLysLysAsnSerSerGluSerValAlaRphe 1425
 OY 2857 AGAGATCTCTGGGAAAGAAAGCGTGACATGGTAAAGCTGCAAGACATTTTGAAACGTGT 2916
 Db 1426 SerGluLeu-----AlaAsnGlyGluPheThrLysAlaGluGluGlu----- 1439
 OY 2917 TTTGTATATATACATAAGTCACTCAAGGCAAGAAAGATGATATATATGATGATGCC 2976
 Db 1439 ----- 1439
 2977 CCCAAATTTTAAATTGAGGAAAAATTGGTGGAACCTAAT--AGAGCCAAAGTATGG 3033
 1440 -----GluLysAsnAlaSerAlaIleArgGluAla 1448
 OY 3034 GAAGCCATGAATGTGATATAATATATTTGAGGATTAATCGGAGACCAATCAACACAA 3093
 Db 1449 GluAlaGluLysLeuLysGluGluIleValLysAsp----- 1460
 OY 3094 AGATGATTATGCGGATATAGTATCATACCAATGGATGATTATATCCACAAATAATTA 3153
 Db 1461 -----LeuAspTyrSerAspIleAspAspLysValLys----- 1471
 OY 3154 AGATGATGACCGAATGGCAGAAATGCTACTGCAAGGTGAG--AAAAAGCTAT 3207
 Db 1472 -----LysIleGluLysIleLysAlaArgGluIle 1480
 OY 3208 GATTAATGAGGAGAGAGCTGTAAGAGTAAAGATTAAGATATGCTCAAGCTGTACG 3267
 Db 1481 LeuLysMetLysGluSerAla----- 1487
 OY 3268 AAGAAGATGTGTACAGTTGTACAGATGCACAGAACTTTAATGAATATATATATATA 3327
 Db 1487 ----- 1487
 OY 3328 ATAGGATTTTGAAGAAACAATGAATATATATATGATATATACAAAGATTAATGAA 3387
 Db 1488 LeuThrPheTyrGluGlu-----SerGluLysPheLysGluMet----- 1500
 3388 CAAGCAAAATGTCTAGTATAGTATGATGATGATGAACTTCCAGTATGCAAAATATCAT 3447
 1501 -----CysSerSerHisMetGluAsnAlaLysGluLysLysLys 1514
 OY 3448 ATAGCAGAGATGTTATGAAATTTTGTGCGAATTAATACCAAAATGTTGGGCAAAAGT 3507
 Db 1515 IleGlu-----TyrLeuLysAsnAsnGlyAspGly 1524
 OY 3508 AATAAAGTGTACTAGTAT----- 3528
 Db 1525 GlyLysAlaAsnIleThrAspSerGluMetGluValGluAsnTyrValSerLysAla 1544
 OY 3529 -----GAAAGTCTGTCATTTGCTACTGTTACTTAACACACAGTATGAAT 3567
 Db 1545 GluHisAlaRheHisThrValGluAlaGluValAspLysThrLysAlaRheCysGluSer 1564
 OY 3568 GTTGAGCATATCTCCATGATACAGGAATTTTATGATATGTCACGACAAATAGCTTT 3627
 Db 1565 IleValAlaIleTyrValThrLysMetAspAsnLeuPheAsnGluSerLeuMetLysGluVal 1584
 OY 3628 -----TGTGATGAAAAAAGTATGATGTAAGATTAACAAAAATATGCTTTAGAGAT 3678
 Db 1585 LysValLysCysGluLysLysAsnAsp-----GluAlaGluLysTyrSerAlaLysLeu 1602
 OY 3679 AAACCAAGACCACTGATGTGTGCTGTGTGTAAGAGTATGAAACCGACAGAGGTA 3738

Db 1603 LysPro-----TyrAspGlyArgIleLysAlaArgValSerGluAsnGluArgLysIle 1620
 OY 3739 ---CAGTTAAAGAAAGAAAAAGCGGAGAAAGAAAGATACGATATGAACATGAAT 3795
 Db 1621 SerGluLeuLysGluLysAlaLysValGluLysLys-----GluSerSerGluLeuAsn 1638
 OY 3796 GATATACCTTAAGAAACGATGGAAGAAACAAGTGAAGATTTGTCATCCAAAAAGAAAT 3855
 Db 1639 AspVal-----SerThrLysSerLeuGluGluIleAspAsnCysArgGluIleLeuAsp 1656
 OY 3856 AGTAATGATATCCGATGGCAATGCGGAATATATAATTTAGTGAAGACCCCTGCTGG 3915
 Db 1657 Ser-----ValLeuSerAsnIleGlyArgVal----- 1665
 OY 3916 TGTATCCCCCTAAGACAAAGATTAATCCGATATCTTCTTGCAATGATATGAATA 3975
 Db 1666 -----LysIleAsnAlaLeuGluTyrPheAspSerAlaAspLysSerMet 1680
 OY 3976 AAAAAATTACAA-----TCACAAGTATATTTAAAGAAAGCTTTTCATCAAAATCTGCACGA 4029
 Db 1681 LysSerValLeuProIleSerGluLeuGluLysAlaGluLysSerLeuAspLysValLysAla 1700
 OY 4030 CGAGAAACATCTCTCATGATATATATATAAGTAAAGATGCGTGAAGCAATGAACCTC 4089
 Db 1701 AlaLysGlu-----SerTyrGluLysAsnLeuGluThrValGluAsnGluMet 1716
 OY 4090 GATTA--GAATTAAAGAAAGCAAAATCTCCCGCATTTTGAGATCCATGCTGTAC 4146
 Db 1717 SerAlaIleAsnValGluGluGlySerLeu----- 1726
 OY 4147 ACATTTGAGATTAATAGATTTTATTTGGAACAGATATATCAAAAGCT----- 4197
 Db 1727 -----ThrAspIleAspLysLysIleThrAsp 1735
 OY 4198 ---CATGCTAGGAGAGTAAACTAAAGACGA-----ATAGATTCCTCT 4239
 Db 1736 IleGluAsnAspLeuLysLysMetLysGluLeuTyrGluGluLysLeuGluLysIle 1755
 OY 4240 TTCAAAAATGCTGACCAAAAAATCTCTAATGGAAGAAACAGCCCAAGAAATGTTGACAGAA 4299
 Db 1756 LysGluAsnAlaAspLysValGly-----LysSerAsnPheGluLeuValGlySer 1771
 OY 4300 CATATGATGATGATATGGAAGCTATGCTATGCTACTATGAATAATTTGGGCAAAAAA 4359
 Db 1772 GluIleAsnAlaLeuLeuAspProSerThrSerIlePheIleLysLeuLysGlu 1791
 OY 4360 GATGATTTTACG-----GAAACTACGCT----- 4383
 Db 1792 TyrAspMetThrGluLysLeuLysLysAsnTyrGluValLysMetAsnGluIleAsnGlu 1811
 OY 4384 -----TACAAACAAGTCAAA-----TTTGTGCAAAAGCACAC 4416
 Db 1812 PheThrLysSerTyrAsnLeuIleGluThrHisLeuSerAsnAlaThrAspTyrSerVal 1831
 OY 4417 ACTTTGAGCAATTTGCCAAAGCAACCCGATTTTATGATGCTTAACGATGTAAGCAG 4476
 Db 1832 ThrPheGlu-----LysAlaGluSerLeuArgGluLeuAlaGluLysGluGlu 1847
 OY 4477 GACTATTTGCTATACACAGCA----- 4497
 Db 1848 GluHisLeuAlaGluArgGluGluGluAlaIlePheLeuLeuAsnAspIleLysLysVal 1867
 OY 4498 -----AATATTTTGAAGATGTCGAGGAAAAATGTAAGTCAAT----- 4536
 Db 1868 GluSerLeuLysLeuLeuLysGluMetLysLysValSerAlaGluTyrGluGluMet 1887
 OY 4537 -----GACCAA 4542
 Db 1888 LysArgAspHisThrSerValSerGluLeuValGluAspMetLysThrIleValAspGlu 1907
 OY 4543 TTGAAGTGT-----GATACAGAAATGTAATTAAGAAATGCGAGGACTGATTAATAT 4593

D 1908 LeuLysThrLeuAsnAspIleSerGluCysSerSerValLeuAsnAsnValSerIle 1927
Q 4594 ATGAAAAAAGAGAGTGGATTCACAAAGATAATATATACAGATGACCGACAA 4653
D 1928 ValLysLysValLysGlu-----SerLysHisIleAspTyrArgArgAspAla 1943
Q 4654 AAAAGATTGCATGACACACATGGTGAATGTTACACATATACTGGAAGAAATGCA 4713
D 1944 AsnSerMetCysGluSerMetVal-----ThrLeuAla 1954
Q 4714 ACAGATTACTTGAACAAATTTACTGCTAGTTGTGTGATAGCCTGGAAGTGC--- 4770
D 1955 AsnTyrPheLeuSerAspGluAlaLysIleSerSerGlyMetGluPheAsnAlaGluMet 1974
Q 4771 -----TCTGTGTACAAAGAATATA 4791
D 1975 LysSerAsnPheLysThrAspLeuGluLeuGluIlePheSerValIleSerAsnSerAsn 1994
Q 4792 CAATTGTTGAAAACAGCGTTCTACTGATGCCGACAAACATGTGGGTGCACAAATTT 4851
D 1995 GluLeuLeuLysLys----- 1999
Q 4852 ATTGAAATATGCACGCAAAATTTACTAATTTCTGATTAAGATAGTCAAGATTTAGTA 4911
D 2000 IleGluGlnAspSerAsnAspVal-----IleGlnLysGlnArgGluSerGluGlnLeuAla 2018
Q 4912 AAGAGGCAAAACACAGCTGCTATTAAAGTGCAAAACAGCTCTAATTAACAATAC 4971
D 2019 LysAspAlaThr-----AspIleTyrAsn--- 2026
Q 4972 TTGAAAGAAATGACTGAAGATGTGCTTTTCCTCTCGTACACTAGTATATTTTCAT 5031
D 2026 ----- 2026
Q 5032 GCATTGATGCAATTATACAGATCCAGAGTAAGATGAATAAGTGGTTCGAAAGAA 5091
D 2027 -----ValIleLysLeuLysAsnGluPheAsnGlnLys 2037
Q 5092 TTGATGGAAGTGGCGGACACGAGGTTACAAATTTGGGTCAATACTCAAGAAAAA 5151
D 2038 LeuGluGlnLys-----LysAsnLysGln 2045
Q 5152 GAAAAAGAAATAAAGCTGCGATGCGCACAAATATTTCTATGAGTCCCGCCCTGT 5211
D 2046 GluValValSerGluLysValArgGluAlaLeuLysArgLeuSerGlnValIleGluGlyIle 2065
Q 5212 AGTCTATGAATATGATTTTATGATTTAAGATATTAATTTAGCTATGATTAATTTG 5271
D 2066 ArgCysHisPheGlnAsnPheHisArgLeuLeuAspAsnThrGluGluLeuGlnAsnLeu 2085
Q 5272 GAA-----GATGAAAAACAAAGACCGACGAAAAATTTGAAGAA 5310
D 2086 LysLysMetValThrIleTyrArgAspLysLysSerGluArgLysSerGluGln 2105
Q 5311 ATTTTAACAAAAATGGAATCATGTTGGCAAGAAAGATGATGACTACAGAAATCCC 5370
D 2106 MetGluAsnGluMetAsnThrTyr-----SerAsnSerIleThrGlnLeuGln 2121
Q 5371 GGTAGTACTGCGGCAAAATTTTCTGAAACGAAATTAAGCAATGCTGGAACCGCAATG 5430
D 2122 GlyIleValValSerAla-----GlyGluSerLysLysLys 2134
Q 5431 ATATGCGGGTCAACAGCTGTAGGATGATGAATAGTGAATAGTGCACAGAGTGAAT 5490
D 2135 -----GluLysLeuGluArgSerAsn 2141
Q 5491 GAAGATCTAAAA-----AAATGTGCTCTGACTTCA----- 5523
D 2142 GluGluMetArgAsnIleSerGluLysIleSerThrIleAspSerLysValIleGluMet 2161
Q 5524 -----GATGATGATTAATCTTANGGGGAAAAATGCGATGAAGGCTACTGCTAT 5571
D 2162 AsnSerThrIleAspGluLeuTyrLysLeuGlyLysAsnGln----- 2176

Q 5572 CAGTTTCTTGATGCTTGGCCGATGGGTGGAAGATTTTGCAAACTATAAGAAAGAA 5631
D 2177 -----AlaHisTrpIleSerLeuIleSerTyrThrAlaAsnMetLys 2190
Q 5632 TTGGAGAAATTTGGAGGGCGGTGAATGATTATACCTGTGTGATATGAAGATTAAGA 5691
D 2191 ThrSerLysLysLeuIleMetLys-----LysGlu 2201
Q 5692 AAGAAATGTACAGATGGCTGTACACATATATAAATTTATAGTGAAGGAAACACAG 5751
D 2202 LysGluAsnThrGluLysCysValAspTyrIleLysAspAsnSerSerSerThrAspGly 2221
Q 5752 TATGAAAAACAAATCAAAAAATTTGTGACAAATTAAGACAAATATATTCGAGCATCT 5811
D 2222 TyrValGluThrLeuLysGlyPheTyrGlySerLysLeuThrPheSerSerAlaSerGlu 2241
Q 5812 GTGGCAAAAGATGACAGG-----GAGCGTCCGAAATATTAGCAAAACATTA 5859
D 2242 IleValGlnAsnAlaAspThrTyrSerValAsnPheAlaLysHisGluLysGluSerLeu 2261
Q 5860 AAAAAATTTGTGAAAAATAAAGTGCAGATTGCAATATATAGTATGAAGAAGTGTCC 5919
D 2262 AsnAlaIleArgAspIleLysLysGluLeuTyrLeuPheHisGlnAsnSerAspIleSer 2281
Q 5920 ACACAGCATTAACCTGATGTGTAATAGTCAAAATATGCCCA-----TCATTAGAC 5970
D 2282 -----IleValGluGlyGlyValGlnAsnMetLeuAlaLeuTyrAspLysLeuAsn 2298
Q 5971 GATGAACCAAAAGAGTGAAGAAAGTGAATTTGTCAATTTGCCACAGAGTCCACACGT 6030
D 2299 GluGluLysArgGluMetAsp-----Glu 2306
Q 6031 GTACGAAGGGAACACCGTACACACAGGATATCACTGATATCAAAACGACGCGATCGAA 6090
D 2307 LeuTyrArgAsnIleSerGluThrLysLeuLysGluMetGluHisSerThrAspValPhe 2326
Q 6091 AAAGAACGAAAACAGCGCCCTACAAACACGCGGAAAAAGTGGAAATCTAACACA 6150
D 2327 Lys-----PrometIleGluLeuHisLysGlyMetAsn----- 2337
Q 6151 GAATGCGACACAAACACGACACCGGACGACGACACACAAACGAAACGAACTCA 6210
D 2338 -----GluThrAsnAsnLysSerLeuGluLysGluLysLysLysSer 2353
Q 6211 ACAGCAACAACACAGATCTGACGTGGCCACATGCTGAAGGCCATCTTGCATTAATA 6270
D 2354 ValAsnAspHisMetHisSerMetGluAlaGluMetIleLysAsnGlyLeuLysTyrThr 2373
Q 6271 CCAGATAGCAGGGGTGGAATAGAG----- 6294
D 2374 ProGluSerValGlnAsnIleAsnAsnIleTyrSerValIleGluAlaGluValLysThr 2393
Q 6295 -----GGTTGTAATCCAAAAACGATGACAAATATCCATAA 6330
D 2394 LeuGluGluIleAspArgAspTyrGlyAspAsnTyrGlnIleValGluGluHisLysLys 2413
Q 6331 ---TGGGTGTATTTGATGATGCTC----- 6354
D 2414 GlnPheSerIleLeuIleAspArgThrAsnAlaLeuMetAspAspIleGluIlePheLys 2433
Q 6355 AAAGAAATGAAAAATGCAATATGATGCTTCCTTAGGAAAAAAATTTATATTAATAT 6414
D 2434 LysGluAsnAsn-----TyrAsn 2439
Q 6415 ATACAATATTTAATTTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 6474
D 2440 LeuMetGluValAsnThrGluThrIleHisArgValAsnAspTyrIleGluLysIleThr 2459
Q 6475 ATTAATGTGCACCAATAGAACTCAATTTTGTGTTAAATATATATATTAATGAATCCT 6534
D 2460 AsnLysLeuValGlnAlaLysThrGluTyr-----GluGlnIleLeuGluValAsnIle 2476

```

OY 6535 GCACAGAAAAATGAAATTCAGAAAAATGCAATTCAGATGATTTAAAGATATATGAT 6594
Db 2477 LysGlnAsnAspMetLeuGlnAsnIlePheLeuLysValSerIleIleGluTyr 2496
OY 6595 TATACATATGATGATTAAGATATGTTTGGAACTATATTTCTAATGATAAAAA 6654
Db 2497 PheGlnAsnValLysLysLysLysGluSerIleLeuAsnAspLeuTyrGluGluTyr 2516
OY 6655 ATTTAATCTGTAACAATACTGTACACACATTT-----CTCAATGAAAAATAAG 6705
Db 2517 LeuLeuLysIleGluLysIleAspGluIleLysArgAsnValThrGluThrLeuSer 2536
OY 6706 AAAAAACAGATAAAAAAGATGAGAAATACGAAATATTTGGAGAAAAATATAA 6765
Db 2537 SerTyrGluLysAspGlnLysMetSerLysAsnLeuLeuGluLysLysSer 2556
OY 6766 AATTTATTTGGAGAAATGATATATGATTAATCTATCATCTACACAGAAAAAC 6822
Db 2557 LysMetMetAsnTyrThrSerIleTyrGluLeuGluValGluLysAsnArg 2576
OY 6823 GAAAAAGAAAAATAGATATATACCAATGACATGACACCAACAGCCCTTCC 6882
Db 2577 AspAlaLysGlnIleLysAsp-----AspAspThrIleLeuAsnSerVal 2591
OY 6883 CTGTAAGAGTTTGA--AAAAGCCCAATTTTGAATGCTGTACAGAAATGC-- 6933
Db 2592 LeuGluAlaIleIleGlnLysArgLysMetAspAlaIlePheSerGlnMetSerAla 2611
OY 6934 -----GCAGAGAAATTTGTATTAAGAGAGAGAGAGAGAG 6966
Db 2612 AspArgAsnProAsnGluTyrLysSerAlaGluLysTyrMetAsnGluAlaAsnGluIle 2631
OY 6967 TTGTTAAATTTGAGCGCGCTGTAGAGATATAGTGTATGTTAGTATGACGGTAG 7026
Db 2632 IleArgGlnLeuGluValLysLeuArgGluIle-----GlyGln 2644
OY 7027 ACACAGAAATGCGAGCGCTGTGTACATATCAAAATTTTATTAAGAGTGAAGAC 7086
Db 2645 LeuValGlnAspSerGluSerIleLeu-----Ser 2654
OY 7087 GAATATGAAAGACAAAGAAAAATTCAAAAAGATTAAGATGCAAAAAAGTTAAGAT 7146
Db 2655 GluMetAsnSerLysLysSerAlaIleGluLysGlnLysThrAlaArgAlaLeuArgTyr 2674
OY 7147 TATCCTTCTCTGAAAGACATAGAGAAAGCAACATGCTCATGAAATTTAAACATG 7206
Db 2675 SerGluAsnAsnArgArgGluGluGluIle-----ArgAla 2686
OY 7207 AAATTAAGAAATATGTCGCAATAGATTTGCTGTATGCAAAAACTTCTTCACAA 7266
Db 2687 ArgValGlnGluMetSerMetAsnAsnAspProThr-----GlnSerGluThrThrHis 2704
OY 7267 CTTCGCAAAAAACACACAAATCACAATCCGATGCTATGATATATGCAAGATCGCTG 7326
Db 2705 SerGluGlySerIleGluLysGluSerAspSerAspGluThrGluLeuThrHis 2724
OY 7327 GATTATGTCCTGAGAAATTAACAAGTGTAGTCTCTGCAACTTTGAAAAAGGATCT 7386
Db 2725 AspAlaGlyAlaAspGluAspSerThrSer-----SerAlaLysGlyAla 2739
OY 7387 AGCATTCATACAAAAAATTAAGTAACCTTAATACCTATGAATTTGTGAGAAAGCA 7446
Db 2740 -----HisGluLeuGluGluGluThrAlaProMet----- 2751
OY 7447 GCATATATTTATTAAGAGAGAAAAATATGATATATACCTTG-- 7494
Db 2752 -----GluGluThrGluMetAsn--AspAsnThrLeuLeuGluTyrAsp 2765
OY 7495 -----AAGAAAAATTTATACCTATGAGCTACAAAGGAA--AAG 7533
Db 2766 ThrThrArgSerAspGluProAspMetHisThrGluAsnThrGlnAspGlyThrTyrGln 2785
OY 7534 GAAAGTAAAAAATAGTGGAGCTAATAATATATCCTTGGATCTTAAGAAACCTTATGACCT 7593

```

```

Db 2786 AsprThrSerAsnSerSerAspGluAlaAspIleLeuAsnGlyLysPheAsnProAsn 2805
OY 7594 GATTAATATATAGCA 7608
Db 2806 ValIleTyrAlaGly 2810

RESULT 3
NEBU_HUMAN
ID NEBU_HUMAN STANDARD; PRT; 6669 AA.
AC P20929; O15346;
DT 01-FEB-1991 (Rel. 17, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nebulin.
GN NEB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95257391; PubMed=7739042;
RA Labelt S., Kolmerer B.;
RT "The complete primary structure of human nebulin and its correlation
RL to muscle structure.";
RN [2]
RP PARTIAL PRELIMINARY SEQUENCE.
RX MEDLINE=88284704; PubMed=3397062;
RA Zeviani M., Darras B.T., Rizzuto R., Salvietti G., Betto R.,
RA Bonilla E., Miranda A.F., Du J., Samlitz C., Dickson G., Walsh F.S.,
RA Dimauro S., Francke U., Schon E.A.;
RT "Cloning and expression of human nebulin cDNAs and assignment of the
RL gene to chromosome 2q31-q32.";
RN [3]
RP GENOMICS 2:249-256(1998).
RN [4]
RP STRUCTURE BY NMR OF 6610-6669.
RX MEDLINE=98179559; PubMed=9514727;
RA Pollou A.S., Millevoi S., Gautel M., Kolmerer B., Pastore A.;
RT "SH3 in muscles: solution structure of the SH3 domain from nebulin.";
RL J. Mol. Biol. 276:189-202(1998).
CC -!- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MAINTAINING
CC THE STRUCTURAL INTEGRITY OF SARCOMERES AND THE MEMBRANE SYSTEM
CC ASSOCIATED WITH THE MYOFIBRILS. BIND AND STABILIZE F-ACTIN.
CC -!- TISSUE SPECIFICITY: MUSCLE-SPECIFIC. LOCATED IN THE THIN FILAMENT
CC OF STRIATED MUSCLE.
CC -!- DISEASE: Defects in NEB are a cause of the autosomal recessive
CC form of nemaline myopathy (NEM2).
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 178 NEBULIN REPEATS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X83957; CAA58788.1; -
DR EMBL; M19668; AAS59916.1; ALT_SEQ.
DR EMBL; M19669; AAS59917.1; ALT_SEQ.
DR PIR; A29979; A29979.
DR PIR; B29979; B29979.
DR PDB; IARK; 28-JAN-98.
DR PDB; IARK; 24-DEC-97.
DR Genew; HGNC:7720; NEB.
DR MIM; 161650; -.
DR InterPro; IPR000900; Nebulin.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.

```

DR pfam: PF00880; Nebulin, 146.
DR PRINTS; PRO0452; SH3DOMAIN.
DR SMART; SM00227; NEBU, 181.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PSS0002; SH3; 1.
KW Actin-binding; Muscle protein; Cytoskeleton; Repeat; 3D-structure.
FT REPEAT 176 107 NEBULIN 1.
FT REPEAT 113 143 NEBULIN 2.
FT REPEAT 148 178 NEBULIN 3.
FT REPEAT 183 213 NEBULIN 4.
FT REPEAT 218 248 NEBULIN 5.
FT REPEAT 253 283 NEBULIN 6.
FT REPEAT 289 318 NEBULIN 7.
FT REPEAT 324 354 NEBULIN 8.
FT REPEAT 363 393 NEBULIN 9.
FT REPEAT 398 428 NEBULIN 10.
FT REPEAT 434 464 NEBULIN 11.
FT REPEAT 502 532 NEBULIN 12.
FT REPEAT 537 567 NEBULIN 13.
FT REPEAT 573 603 NEBULIN 14.
FT REPEAT 611 641 NEBULIN 15.
FT REPEAT 711 741 NEBULIN 16.
FT REPEAT 749 779 NEBULIN 17.
FT REPEAT 820 850 NEBULIN 18.
FT REPEAT 858 888 NEBULIN 19.
FT REPEAT 893 923 NEBULIN 20.
FT REPEAT 924 954 NEBULIN 21.
FT REPEAT 959 990 NEBULIN 22.
FT REPEAT 993 1023 NEBULIN 23.
FT REPEAT 1028 1058 NEBULIN 24.
FT REPEAT 1064 1094 NEBULIN 25.
FT REPEAT 1102 1132 NEBULIN 26.
FT REPEAT 1137 1167 NEBULIN 27.
FT REPEAT 1168 1198 NEBULIN 28.
FT REPEAT 1204 1234 NEBULIN 29.
FT REPEAT 1237 1267 NEBULIN 30.
FT REPEAT 1272 1302 NEBULIN 31.
FT REPEAT 1308 1338 NEBULIN 32.
FT REPEAT 1346 1376 NEBULIN 33.
FT REPEAT 1381 1411 NEBULIN 34.
FT REPEAT 1412 1442 NEBULIN 35.
FT REPEAT 1448 1478 NEBULIN 36.
FT REPEAT 1481 1511 NEBULIN 37.
FT REPEAT 1516 1546 NEBULIN 38.
FT REPEAT 1552 1582 NEBULIN 39.
FT REPEAT 1590 1620 NEBULIN 40.
FT REPEAT 1625 1655 NEBULIN 41.
FT REPEAT 1656 1686 NEBULIN 42.
FT REPEAT 1692 1722 NEBULIN 43.
FT REPEAT 1725 1755 NEBULIN 44.
FT REPEAT 1760 1790 NEBULIN 45.
FT REPEAT 1796 1826 NEBULIN 46.
FT REPEAT 1834 1864 NEBULIN 47.
FT REPEAT 1869 1899 NEBULIN 48.
FT REPEAT 1900 1930 NEBULIN 49.
FT REPEAT 1936 1966 NEBULIN 50.
FT REPEAT 1969 1999 NEBULIN 51.
FT REPEAT 2004 2034 NEBULIN 52.
FT REPEAT 2040 2070 NEBULIN 53.
FT REPEAT 2078 2108 NEBULIN 54.
FT REPEAT 2113 2143 NEBULIN 55.
FT REPEAT 2144 2174 NEBULIN 56.
FT REPEAT 2180 2210 NEBULIN 57.
FT REPEAT 2213 2243 NEBULIN 58.
FT REPEAT 2248 2278 NEBULIN 59.
FT REPEAT 2284 2314 NEBULIN 60.
FT REPEAT 2322 2352 NEBULIN 61.
FT REPEAT 2357 2387 NEBULIN 62.
FT REPEAT 2388 2418 NEBULIN 63.
FT REPEAT 2423 2453 NEBULIN 64.
FT REPEAT 2456 2486 NEBULIN 65.
FT REPEAT 2456 2486 NEBULIN 66.

FT REPEAT 2491 2521 NEBULIN 67.
FT REPEAT 2527 2557 NEBULIN 68.
FT REPEAT 2565 2595 NEBULIN 69.
FT REPEAT 2600 2630 NEBULIN 70.
FT REPEAT 2631 2661 NEBULIN 71.
FT REPEAT 2666 2696 NEBULIN 72.
FT REPEAT 2699 2729 NEBULIN 73.
FT REPEAT 2734 2764 NEBULIN 74.
FT REPEAT 2770 2800 NEBULIN 75.
FT REPEAT 2808 2838 NEBULIN 76.
FT REPEAT 2843 2873 NEBULIN 77.
FT REPEAT 2874 2904 NEBULIN 78.
FT REPEAT 2909 2939 NEBULIN 79.
FT REPEAT 2942 2972 NEBULIN 80.
FT REPEAT 2977 3007 NEBULIN 81.
FT REPEAT 3013 3043 NEBULIN 82.
FT REPEAT 3051 3081 NEBULIN 83.
FT REPEAT 3086 3116 NEBULIN 84.
FT REPEAT 3117 3147 NEBULIN 85.
FT REPEAT 3152 3182 NEBULIN 86.
FT REPEAT 3185 3215 NEBULIN 87.
FT REPEAT 3220 3250 NEBULIN 88.
FT REPEAT 3256 3286 NEBULIN 89.
FT REPEAT 3294 3324 NEBULIN 90.
FT REPEAT 3329 3359 NEBULIN 91.
FT REPEAT 3360 3390 NEBULIN 92.
FT REPEAT 3395 3425 NEBULIN 93.
FT REPEAT 3428 3458 NEBULIN 94.
FT REPEAT 3463 3493 NEBULIN 95.
FT REPEAT 3499 3529 NEBULIN 96.
FT REPEAT 3537 3567 NEBULIN 97.
FT REPEAT 3572 3602 NEBULIN 98.
FT REPEAT 3603 3633 NEBULIN 99.
FT REPEAT 3638 3668 NEBULIN 100.
FT REPEAT 3671 3701 NEBULIN 101.
FT REPEAT 3706 3736 NEBULIN 102.
FT REPEAT 3742 3772 NEBULIN 103.
FT REPEAT 3780 3810 NEBULIN 104.
FT REPEAT 3815 3845 NEBULIN 105.
FT REPEAT 3846 3876 NEBULIN 106.
FT REPEAT 3914 3944 NEBULIN 107.
FT REPEAT 3949 3979 NEBULIN 108.
FT REPEAT 3984 4014 NEBULIN 109.
FT REPEAT 4021 4052 NEBULIN 110.
FT REPEAT 4057 4087 NEBULIN 111.
FT REPEAT 4088 4118 NEBULIN 112.
FT REPEAT 4123 4153 NEBULIN 113.
FT REPEAT 4156 4186 NEBULIN 114.
FT REPEAT 4191 4220 NEBULIN 115.
FT REPEAT 4226 4256 NEBULIN 116.
FT REPEAT 4264 4294 NEBULIN 117.
FT REPEAT 4299 4329 NEBULIN 118.
FT REPEAT 4330 4360 NEBULIN 119.
FT REPEAT 4365 4395 NEBULIN 120.
FT REPEAT 4400 4430 NEBULIN 121.
FT REPEAT 4435 4465 NEBULIN 122.
FT REPEAT 4471 4501 NEBULIN 123.
FT REPEAT 4544 4574 NEBULIN 124.
FT REPEAT 4575 4605 NEBULIN 125.
FT REPEAT 4610 4640 NEBULIN 126.
FT REPEAT 4645 4675 NEBULIN 127.
FT REPEAT 4680 4710 NEBULIN 128.
FT REPEAT 4716 4746 NEBULIN 129.

Alignment Scores:

Pred. No.: 8, 41e-13
Score: 420.50
Percent Similarity: 30.948
Best Local Similarity: 17.838
Query Match: 2.168
DB: 1
Matches: 6669
Conservative: 760
Mismatch: 1439
Indels: 1505
Gaps: 197

US-10-087-013-1 (1-10628) x NEBU_HUMAN (1-6669)


```
OY 81 TATAAAGAAAGTCACAAAAGTCAG---AAATGTTT-----GGAAGCTATGCCAA 131
||||| : : : : : ||| ||||| : : : : : |||
Db 705 TATLTSALAGLUTGCLUGLULASPRYSGLYSCTYTRPHEPROGINTHILIEHGIN 724
OY 132 AAATTAAGACATCCATCAAAATA-----TGCAAAAGACATGCGATTCGTGGAAGG 185
||||| : : : : : ||||| : : : : : |||
Db 725 GUTYTRSPRALILIELYLSLEUASPRINCYS-LYSASPHISHTYTRYLSVALHISPR 744
OY 186 GGATTTGACGAAGACGAGATTCGTGTCCTTCCTGCGCAGCAATTAAGCAATA 245
||||| : : : : : ||| ||||| : : : : : |||
Db 744 OASPRYSTRYLS-----PHEITHALVALTHASPERVAL----- 757
OY 246 TTATATCCATATCCATTAATTTAGATCATAGAACACATCTAATTTACGGTAT----- 300
||||| : : : : : ||||| : : : : : |||
Db 758 -----LeuLeuGlnAGLInLeuAsnThrLysGlnLeuSeraspLeuAsnThrLysAL 775
OY 301 -GATGATGTGAATTTGAGACATCTTGCCAT----- 330
||||| : : : : : |||
Db 775 ALYSHISGLUGLULUATGRHELYSCYSHISILEPROALASPALAPROGInPHEILEGI 795
OY 331 -----GGTAGAACAACAAACCGATTTGATGAGATGAGATGAGATGCA 371
||||| : : : : : ||| ||||| : : : : : |||
Db 795 nHISATGVALASnLALUTYRAsnLeuSerAspAsnVALTYRlySGInSPTRPRGLYLS-- 814
OY 372 ATGTGAAATTAATAGTATTTATAAAGAAAATGATGCTATGACCTGTGCGCACCC 431
||||| : : : : : ||||| : : : : : |||
Db 815 -----SerLYSALALYSLSYPHeAsPILeLYSVALALIEPROLeuALAL 832
OY 432 TAGAAGACGACATGTGTGTATAAACTTGGAAGCTCTAATGATTAATATCCCAAA 431
||| : : : : : ||| ||||| : : : : : |||
Db 832 ALYS----- 838
OY 492 TATTCATGATTTATGGGAAATGACTAGTTACAGCAAAATACGAA----- 537
||||| : : : : : ||||| : : : : : |||
Db 838 nTHSerASP-----VALMETYTRYLSYASPTYRGLULYSerLYSGLY 854
OY 538 -----GGTGAATCAATGTGTATATATCATCCACATTAAGGAACCTTCAGACCTGTAC 590
||||| : : : : : ||||| : : : : : |||
Db 854 smETILEGLYALeUSeRILEAsnSPAsPRLYs----- 866
OY 591 TGTCTTTCGACGAAGTTTTCACAGATAGGTGATTAATGTAAGGAATAGATGTTTAA 650
||||| : : : : : ||||| : : : : : |||
Db 866 ----- 866
OY 651 ACCAAATGTCATGACAAAGTAGAAACGGGTCCGAGAGCTTTTCAAGAAATACATCA 710
||||| : : : : : ||||| : : : : : |||
Db 867 -----MetLeuHISSerLeuLYSTHIALALYSAs 876
OY 711 TCGAATGGAAGATGAGTAAAAATGATTACATCTGATGATCGAATTAATATATA 770
||||| : : : : : ||||| : : : : : |||
Db 876 nGInSerASPARGLUTYRArgLYSAsPTRY-----GLYSSerLYSTHILETYTH 894
OY 771 ATTAAGAGAAGCATGTGGAATGTGAATTAAGTAATGAAGCTATTAACATGTGA 830
||||| : : : : : ||||| : : : : : |||
Db 894 rALAPROLeuAsPMeTLeuGlnVALThrGlnALALYSLSerGlnALALIEALISerAs 914
OY 831 TCGCATCATTAATCTGATATTGTCATCAAGACAGAAATATACACCTATTATTGCAA 890
||||| : : : : : ||||| : : : : : |||
Db 914 rPALAsPTRYLS-----HISILeUHisSerLYSTYRPROAsPSeRILEAs 932
OY 891 TCTTAATAGCGCCATTAACAAGAAAGGTCCTACCAATTTAGATTATGTCCTCAATA 950
||||| : : : : : ||||| : : : : : |||
Db 932 nVALSPLeuALALYSALATYRALeUInSerASPALeGLUTYRlySALAsPTRY 952
OY 951 TTTAGTGTGCTTC----- 963
||||| : : : : : |||
Db 952 rASnSerTPMeTLYSGLYCYSGLYTRYPVALPROBHeGLYSerLeuGluMeTGLYLSAL 972
OY 964 -----GACGATGGCGAGAGAGCTTTTCCGAAAGAAAGAAATATTAAATGAA 1010
||||| : : : : : ||||| : : : : : |||
Db 972 ALYSATrGALISerAsPILeUAsnGLULYSLYSTYRArgLInHISPROAsPThrLeuLY 992
OY 1011 A-----AAGTCAGAGACTCGTCG 1031
||||| : : : : : ||||| : : : : : |||
Db 992 sPHeThSerILEGLULASPALAPROILeThrVALGInSerLYSILEAsnGlnALeGInTr 1012
OY 1032 TAATGAC----- 1038
||||| : : : : : ||||| : : : : : |||
Db 1012 gSeRAsPILeALeTLYLSALALYSGLUGLULIeILeHISAsnTYRAsnLeuPROPr 1032
OY 1039 -----AAGAACGCTTATATTGTAGTCATATGACATGATTTGACACAATATTGG 1091
||||| : : : : : ||||| : : : : : |||
Db 1032 OASPRLeuPROGInPHEILEGlnALALYSVALASnALeTYRAsnILeSeRGLAsnMeTY 1052
OY 1092 GAAAAAAGATTTTTCATTTGGATTAATAGCTACTGCTGTGCAATTAATGCAAGT 1151
||||| : : : : : ||||| : : : : : |||
Db 1052 rLYSALASPRLeuLYASPRLeuSerLYSGLYTRYAsPLeuAVThrAsPALALIEPR 1072
OY 1152 TTTTGAAGTTTGGTTAGGAATCAACAGACCA-----TTTAAAAAACAAAA 1199
||||| : : : : : ||||| : : : : : |||
Db 1072 OILeATrGALALALeLYSALALeALeAGLInLALeALeASerASPVALGInTYTRYLSYASPTY 1092
OY 1200 AGAAAAATATGAAAAAAGAAATACATCATAT-----TTATGCAACGATTAACAAATTTGT 1253
||||| : : : : : ||||| : : : : : |||
Db 1092 rGLYLSALALYSGLYLYSMeTVALGLYPHeGInSerLeuGlnAsPAsPRLYsLeuVA 1112
OY 1254 CAATATATTATAT-----AGTGAATATTATAACAATTTTATGAAA 1295
||||| : : : : : ||||| : : : : : |||
Db 1112 IHSTYrMeTAsnVALALALeLYSILEGInSerASPARGLUTYTRYLSYASPTYGLULY 1132
OY 1296 ACTTAAGCAACGACATATGCACTAATGACACTTTTAAATATACATAATGAAAGAA 1355
||||| : : : : : ||||| : : : : : |||
Db 1132 sTHrLYs-----SerLYSTYRAsnThrPROHISAsPMeThEAsnVALALALALeLYS 1151
OY 1356 GTATGTGAAGAGATTAACAGCAAGAAAGATTTACTTTACTAACAGCTGTGATGA 1415
||||| : : : : : ||||| : : : : : |||
Db 1151 S-----ALeGlnAsPAL-----VALSerAsnVALAsnTY 1161
OY 1416 CAAGAGATTTATTCGTCAGATATTGCCAATGTCGCCAGTCGGCGGCAATG 1475
||||| : : : : : ||||| : : : : : |||
Db 1161 rLYSHISSerLeuHISHTYTRYThrLYLeuPROAsPRLAMeTAsPRLeUGInLeuSerLY 1181
OY 1476 TGATGTATTAATATACACACAAATCGATTAATGATCGTGAACGTGAATAATGAGA 1535
||||| : : : : : ||||| : : : : : |||
Db 1181 sAsnMeTMeTGLInLE-----GInSerASPAsn-----VALTYRlySGLUAs 1195
OY 1536 CTATTAACCTCCATG-----GGTGAAGCTACTAATTAATCACTGCTTTA 1583
||||| : : : : : ||||| : : : : : |||
Db 1195 PTRYAsn-----AsnThrPMeTLYSGLYILEGTYTRPILePROILeGLYSerLeuAsPVALGI 1214
OY 1584 TAGTGTAATGACAAAGGTATAT-----AC 1610
||||| : : : : : ||||| : : : : : |||
Db 1214 ULYSVALYSLSYsALeGLYAsPALALeUAsnGLULYSLYSTYRArgLInHISPROAsPTh 1234
OY 1611 ACAAATAATTAGAAATTTTGTACAGCTCACTACATTAATTAAGATTAATAATATCAAA 1670
||||| : : : : : ||||| : : : : : |||
Db 1234 rLeuLYSPHeThSerILEVALAsPSeRPROVALMeTVALeGlnALALeYSGLInAsnThrLY 1254
OY 1671 ATGG-----GAATGCTATTATAGCATGAAATATTAATAGATGTAAGTGAACAAAA 1724
||||| : : : : : ||||| : : : : : |||
Db 1254 sGlnVALASerAsPILeUeUTYRlySALALYSGLYGLUAsP-----VALYSHISLY 1271
OY 1725 TACTGAATCAATATATATCTTAAGATTAATATCATTTTCAT-----AATTTTTGAAAT 1781
||||| : : : : : ||||| : : : : : |||
Db 1271 sTYThrMeTSeRPROAsPLeuPROGInPHEUGlnALALeLYCYAsnALeALeTYSerIL 1291
OY 1782 ATGGCTTACATATTATTAAGGCTACTATTAAGTGAATGACAAACTTAACCTGTAT 1841
||||| : : : : : ||||| : : : : : |||
Db 1291 eSeRAsPVALCYSTYTRYLSYARGSP-----TRPHISAsPLeuILeARGLYSGLY-- 1307
OY 1842 AAATATATACACACGCAATTTGATGATGAATGTAACAGAAATTCGTATGTTTGACAG 1901
||||| : : : : : ||||| : : : : : |||
Db 1308 -AsnAsnVALeUGLYAsPALALIEPROILeThrALALeLYSALA----- 1322
OY 1902 ATGGGTTAAACAAAAGAGAAAGATGGAATGATTAAGAAAGAACTGTTCACAAAAAAA 1961
```

```
Db 1323 -----Serat 1324
OY 1962 GAATATACAGCAATCGATATAGTAAATATTAATCTTTTGAAGGTATTTTAA 2021
Db 1324 gaaTlleaJaSerapTyr-----1330
OY 2022 AGTTATGATTAACCTTGACAAAGATGAGCAAAATGAGAAAGACTTATGAAATATAA 2081
Db 1331 -----LysTyrLysGluAlaTyrGluLysSerLys 1340
OY 2082 AAGAAAAAATATGAGTT-----2100
Db 1340 sgLyLysHISvalIGlyPheArgSerLeuGlnAspProLysLeuValHisTyrMetAs 1360
OY 2101 -----TCCAAATTTGGAATAATATAGGACATTTAGAGAAATGAACTCTGTAGGA 2156
Db 1360 nvalAlaLysLeuGlnSerAspArgLysLysAsn-----TyrGI 1375
OY 2157 TCACCTTAAAGAAACGCGACGATATGTAAAGCAATTAATACAAAGCAAGCATGTGAAC 2216
OY 1375 uaaNTThrLysThrSerTyrHisThrProGlyAspMetValThrIleThrAlaLysMe 1395
OY 2217 ATCCCAATTAATGCAACAACACCCCTGTCTTAACCTCGTGAGCGACGCAACCC-----2271
Db 1395 lAlaGlnAspValAlaThrAsnValAsnTyrLys-----GlnProLeuH1 1410
OY 2272 -----ACTAAATAATTAAGAAAT 2291
Db 1410 sHISlyrThrTyrLeuProAspAlaMetSerLeuGlnHisThrArgAsnValAsnGlnI1 1430
OY 2292 AGCACAATACTTTAAAGAGTCATACGAGAGCAAGAAAT-----CGTGCTCT 2342
Db 1430 eGln-----SerAspAsnValTyrLysAspGluTyrAsnSerPheLeuLysGlyI1 1447
OY 2343 T-----CAATAATTAAAGAAAGCAGACAGAGG 2372
Db 1447 eGlyTrrPleProIleGlySerLeuGlnValGluLysValLys-----LysAlaGlyAspAl 1466
OY 2373 TATATTAACGTCGGGGTGGAGAG-----AAGCACTTCAAG 2411
Db 1466 alAuAsnGluArgLysTyrArgGlnHisProAspThrValLysPheHisSerValProAs 1486
OY 2412 CAATTTATGTAGAAATAATGATTAACCAATTCGTAATCTTGTTTCAATGAGACC 2471
Db 1486 pSerMetGlyMetMetLeuAlaGlnHisAsnThrLysGlnLeuSerAspLeuAsnTyrLys 1506
OY 2472 ATGTGANTGCAAAAGCGACAGGTGATGTATACAAACAAGATTGTCTGACAGACTGAATG 2531
Db 1506 svaIGluGlyGlu-----LysLeuLysHISLysTyr 1516
OY 2532 GGAAGTGCATCGGAGACACATGCGTAAAGATCAGAGATGTATTTATGCTCCTAGAG 2591
Db 1516 rThrIleAspProGlu-----LeuPro-----1523
OY 2592 ACAGCATATATGTACATCCAAATTTGCAACATTACAAACCGATGATCACCACCTTAATGC 2651
Db 1524 -----GlnPheIleGlnAlaLysValAsnAlaLeuAsn-- 1534
OY 2652 TAATATTTGATGATTTACTTAATAATCTTTTGGGGAGTGTCTGTATCAGAA 2711
Db 1535 -----MetSerAspAlaH1 1539
OY 2712 ATATGAGCAAAAC-----AAGATAATACGAATGTATAAGAAAGAAATACCTAAAGG 2765
Db 1539 sTyrLysAlaAspTrrPyrLysThrIleArg-----LysGlyTyrTrsPleuArgProAs 1557
OY 2766 CCCCAGAAAGTACGACCCCAAAACCCAGACAACTATGTCTGAGCTTATAGCTTACAG 2825
Db 1557 palAlleProIleValAlaLysSerSerArgAsnIleAlaSerAspCysLysTyrLys 1577
OY 2826 TTTTGAGATATAGGTATTAATTCGAGAGAAGATCTCTGGAAAGAAACGCTGACAT 2885
Db 1577 sgLuala-----1579
OY 2886 GGTAAAGCTCAGAGACATTTGGAAACTGTTTGGTAATATACATAAGCTCAAGG 2945
Db 1580 -----TyrGluLysAlaLysGI 1585
OY 2946 CAAA-----GGAATGATTAATATATGATGATGCCCCCAATTAATTAATTTGAGGGA 2999
Db 1585 yLysGlnValIGlyPheLeuSerLeuGlnAspAsp-----ProLysLeuValHisTyr-----1602
OY 3000 AAATTTGCTGGGAAGCTTAATAGACCAAAAGTATGGAAGCCATGAATGTATTAATA 3059
Db 1603 -----MetAsnValAlaLysIle-----GlnSerAspArgGluTyr 1614
OY 3060 TTTCGAAGATTAATCGGAGACCAATCAACAAAGATTAATTTGGGATATAGTATCA 3119
Db 1614 T-----LysLysGlyTyrGluAlaSerLysThrLysTyr-----H1 1626
OY 3120 TACACCATTTGGAT-----3132
Db 1626 sThrProLeuAspMetValSerValThrAlaAlaLysLysSerGlnGluValAlaThrAs 1646
OY 3133 -----GATTATATCCCAAAAAATTAAGA-----3156
Db 1646 nAlaAsnTyrArgGlnSerTyrHisTyrThrLeuLeuProAspAlaLeuAsnValGI 1666
OY 3157 -----TGATGACGCAATGCGCAGA 3176
Db 1666 uHISserArgAsnAlaMetGlnIleGlnSerAspAsnLeuTyrLysSerAspPheThrAs 1686
OY 3177 ATGC-----TACTGCAAGTGCAGAAAAAGCACTATCAATAGTTGAAGA 3221
Db 1686 nTrrPmetLysGlyIleGlyTrrPvalProIleGlnSerLeuGlnValGluLysAlaLysLys 1706
OY 3222 GAAGTGTAAAGAGCTTAAGATTAAGATTAATGCTCAAGCTGTACGAAAGAGTGTAC 3281
Db 1706 sAlaGlyGlnIleLeuSerIleLysLysTyrArgGlnHisProGluLysLeuLysPheThr 1726
OY 3282 AGCTGTACGAAGTGCACAGACCTTGT-----AATGAATATATGATATAT 3329
Db 1726 rTyrAlaMetAspThrMetIleGlnAlaLeuAsnLysSerAsnLysLeuAsnMetAspLys 1746
OY 3330 AGCATTTAGAAAGAACATGCAAT-----3354
Db 1746 sArgLeuTyrThrGluLysStrAsnLysAspLysThrThrIleHisValMetProAspThr 1766
OY 3354 -----3354
Db 1766 rProAspIleLeuLeuSerArgValAsnGlnIleThrMetSerAspLysLeuTyrLysAl 1786
OY 3354 -----3354
Db 1786 aGlyTrrPgluGluLysLysLysGlyTyrAspLeuArgProAspAlaIleAlaIleLys 1806
OY 3355 -----ATAATATCAGAT--AAATACAAAGATTAATACATGACA 3389
Db 1806 sAlaAlaArgAlaSerArgAspIleAlaSerAspTyrLysTyrLysLysAlaTyrGluGI 1826
OY 3390 AGCACAATGCTGTATGATATAGTGTATGAAGCTTCAGACTCCCAAAATCATAT 3449
Db 1826 nAlaLys-----GlyLysHISL1 1832
OY 3450 A-----GACAGAAATGTATTAATTTTGT-----3474
Db 1832 eGlyPheArgSerLeuGlnAspAspProLysLeuValHisPheMetGlnValAlaLysMe 1852
OY 3475 -----TCGGAATTAATACCAACAAAATGCTGCCAAAAGTAATAAAGTGT-----3519
Db 1852 tGlnSerAspArgGluTyrLysLysGlyTyrGluLysSerLysThrSerPheHisThrPr 1872
OY 3520 -----ACTAGTGAATGAAGGTGTGTACTGTGTACTAAGCACAC 3557
Db 1872 oValAspMetLeuSerValValAlaAlaLysLysSerGlnGluValAlaThrAsnAlaAs 1892
```

[illegible]

OY	4413	CACACATTGG-----GAGGAATTTTCCCAACGACC-----CAGTTTTACG	4455
Db	2195	aethrglwtgylalasearglntstygabgnlnhsrproserhsrphelnlheylsty	2215
OY	4455	ATGGCTAACCGGAAATGTGCAC-----GACTATTGCTATACGACGAAA	4495
Db	2215	s---leuthrrasermetaalvalleualalyuglnasplalnshisthmetsnly	2235
OY	4500	ATAATTGAAGAATGTCGAC---GAAAAATGTAACTCCAATGACCATTGAAGTGATAC	4555
Db	2234	shisheulythrlliesarptrrsalnysarplysthrlensivalmetcrpsarsh	2255
OY	4557	AGAAATTAATAAAGAATCCGACGACSTCGTTAAATATATGMAAAAAAAAAAGCTGAT	4615
Db	2254	r-----Proardlleugln-----Alalyuglnasnclnthle	2266
OY	4617	TCCACAAGATAATTTCACAG-----GATGAACGCGACAAAAAAGTTGATAG	4667
Db	2266	wtyrsertglnysleuytrlysleuglyttrpgnglnlalaeluyltgyltgrasple	2286
OY	4668	A-----CAACACATGTGTGAATGTTACAGACTTACTGGAAACGAATCAACAGATTA	4721
Db	2286	uprovalasprallaleserValglnleualalysrslaserahgsarpllealsesapry	2306
OY	4722	CTTGACACGAAA---TTTAGCTGTAATGTGTGGTAAACCTGGAAGTCCCTA----	4773
Db	2306	rlsystrylsynglnlytgraglysglnleuglnshlsvalglrPhetargserteugl	2326
OY	4774	-----GTGGTACAAAAGAAATATATACAAATTTGTNAGAAAAACAGCTTACTA	4817
Db	2326	nasparsprolyslenuvallneusermetsbnvalalalysemgnglnserglnatggluty	2346
OY	4818	TGATCCGACAAACATTTGTGGGTGCACAAAATTTATTGAAAATGACGACAAATATCTNA	4877
Db	2346	rlsylysasprheglwlustyrplrstlysrpheserperovalaspmelleuglyva	2366
OY	4878	CATTTCGATGAATGAATAGTCGCCAAAGATTAAGTAAAGGACGAAACACAGGTCTATTAA	4937
Db	2366	lvalleualalys---lyscysglnleuleualSerapsval-----	2379
OY	4938	GTGGCAAAACAAAGCTCTTAATACTACATCAATAAC---TGAAAGAAATGAGCT---	4986
Db	2380	-----Aspyrlrysasnylrleuhnsnglntrphrncysleupr	2392
OY	4987	-----GAAATGTGCTTTTTCCTTCCTGCTGCTACSTATATGTGTTTCATGCATTGGA	5039
Db	2392	oasprlnasnasprvalvalalnalalysrvaltyr-----	2404
OY	5040	TGGCAATTAATACAGATCCCAAGTCTTAAAGTGAATGGTTGGCAAAAAGATTGATGGA	5099
Db	2405	-----glutenglnerglusasn---leutyrlsysersapleugl	2417
OY	5100	AGTGGCGGCAACGAGAGGTCACAT---TTGGGTCAATACTCAAAAAMAAAAAAGAAA	5156
Db	2417	utrpleuarlglytlleolytrpserproleunglyserleuglnalaglulysasnlvsar	2437
OY	5157	AGAGAAATAATAAAACGCGGAGTCCGACAAATATTCCTTATGAGTCCCGGCTGTATGTC	5216
Db	2437	galsercglulleleserulnys-----lystyrarglnprprobspargas	2454
OY	5217	TATGAATAATAGTTTTTATGATTTAAGAGATATAATTACTAGATATGATAATTGGAAAG	5276
Db	2454	nlysrPheThrSerlleproasprlametksrlllevalleualalysrthrxsn-----	2471
OY	5277	TGAAAAACAAAACCGACGAGAAAATTTGACAGAAAATATTTAACAAAATGGAACATCATGT	5336
Db	2472	-AlalyasasnaryserAspratydeulytrglnlatrpsrlysasprlystrhcnll	2491
OY	5337	TGGCAAAAGAAAGTAGATG-----ACTACAGAGAAATCCCGGTAGTACTGC	5381
Db	2491	eHsisllemerProasprhnrProasprllevalleualalysalabsnneuileasnthre	2511
OY	5382	GCGAAAAATTTTCTGACACGAAAAATAGAAGATGTGTGGAACGCAATGATATCCGGTA	5441

QY 7134 AAGATTAAGAT-----TACCTTCTACGAAAGAGA 7166
Db 3140 eTYLysSerAspLeuGlnTrpLeuArgGlyIleGlyTrpValProIleGlySerMetAs 3160
QY 7167 CATGAGACAGGCAACATGCTGCTCATGAATATTAAAC----- 7203
Db 3160 pValValLysCysLysArgAlaAlaGluIleLeuSerAspAsnIleTyArgGlnProPr 3180
QY 7204 -----ATGAATTTAAAGATTTATGTGCAATAAGA 7235
Db 3180 cAspLysLeuLysPheThrSerValThrAspSerLeuGlnValLeuAlaLysAsnAs 3200
QY 7236 TTGTTCCTGATGCAAAAACCTTTCACACACTACCAAAAACAACACACATCACAATC 7295
Db 3200 pAlaLeuAsnMetAsnLys-----ArgLeuTyThrGluAlaIleTrpAspLysAsp 3217
QY 7296 ATCCGATGCTAAATATATGCGCAATCGCGATTTATGTCGGAAGAAATTAAACAAGT 7355
Db 3217 sThrGlnValHisIleMetProAspThrProGluIleMetLeuAlaArgGlnAsnLysI 3237
QY 7356 TGAAGTCTCGAAGCTTTCAAAAAGGAGATCTATGATTCATACAAAAA----- 7404
Db 3237 eAsnTySerGlnSerLeuTyArgGlnIleMetGluGluAlaLysGluIleTyTrAs 3257
QY 7405 -ATTACTGAACCTTAATACCTATGATGTGTAGAGAA-----GCACCAATA 7451
Db 3257 pLeuAlaGSerAspAlaIleProIleValAlaAlaLysAlaSerArgAspIleAlaSerAs 3277
QY 7452 TTTATTTATTAAGACGCAAAATTAATATG----- 7482
Db 3277 pTyLysTyLysGluAlaTyArgLysGlnLeuGlnHisIleGlyAlaArgAlaVa 3297
QY 7483 -----GATTTTAC 7490
Db 3297 lHisAspAspProLysIleMetTrpSerLeuHisIleAlaLysValGlnSerAspArgI 3317
QY 7491 CTTGAAGCAAAATTT-----ATACCTATGAGCTTACAA 7526
Db 3317 uTyTyLysLysAspPheGlnLysTyThrArgTySerSerProValAspMetLeuG 3337
QY 7527 GCAAAAGCAAAATTAATATGTTGCACTATTAATATCTTCCGATCCTTAAGAAACCTTA 7586
Db 3337 yLleValLeuAlaLysLysCysGlnThrLeuValSerAspValAspTyLysHisProLe 3357
QY 7587 T-----GCACCTGATAAA-----TATATNGAAGAAAGAAACCTTGGA 7625
Db 3357 uHisGlnCysIleCysLeuProAspGlnAsnAspIleIleHisAlaArgLysAlaTyTrAs 3377
QY 7626 AATATGAGAGAAATCTGTTTAAGTAGATTTATGATGAAATGTAACAAATTTCAA 7685
Db 3377 pLeuGlnSerAspAsnLeuTyLysSerAspLeuGlnTrpMet----- 3391
QY 7686 GTTCTATGACGAGAAAAAGATATGTGTACCTCCAAAGACAGAAACATATGTGCTTAAG 7745
Db 3392 -----LysGlyIleGlyTrpValPro----- 3398
QY 7746 GAATTTAGATGAATTAATTAATGAAGACTTAAGATAGTATATTCTTCTAAATAGT 7805
Db 3399 ----lLeAspSerLeuGlnValValArgAlaLysArgAlaGlyLeuLeuLeuSerAspTh 3417
QY 7806 TCGTGAACTGCACGAATGAAGAAATGACATTAATAAAAACTTCAACTCA----- 7857
Db 3417 rIleTyArgGlnArgProGlnTrpGluThr-----LysPheThrSerIleThrAs 3433
QY 7858 -----GAGAACGGGTCGCAATGCAATATGATATGATCTATGAATATAGTTGCG 7910
Db 3433 pThrProGlnValValLeuAlaLysAsnAsnAlaLeuAsnMetAsnLysArgLeuTyTrTh 3453
QY 7911 TGATCTGGGTGATAGTTAGAGAACAGATATGTTACGAATGGTGGTTACTTACCTCC 7970
Db 3453 rGluAlaIleTrpAsp-----AsnAspLysLysThrIleHisValMetProAspTh 3469
QY 7971 CGTGAATAATAATTAATAGGTTTGTGAATACATATATGAAATGAGAAATTAATA 8030
Db 3469 rProGluIleMetLeuAlaLysLeu----- 3477
QY 8031 TAAAGTAGAATAATTAATACACAT-----GTACAAACGTTTCGTTTC 8072
Db 3478 ----AsnArgGlnAsnTySerAspLysLeuTyLysIleuAlaLeuGlnGlnSerLys 3496
QY 8073 TCGTTGCTGGATGCTTAATAGAAAGATTTTGAAGACATGACGTCAAGACCAAGA 8132
Db 3496 sGluGlyTyTrAsPheLeu-----ArgLeuAspAlaIleProIleGlnAlaAlaLysSer 3515
QY 8133 ACATGCAAAACCTTTTAGAAGAAAGATGAGATGATTTGAACGATTAACATTAACA 8192
Db 3515 gAspIleAlaSerAspTyLysTyLys-----GluGlyTyTrAsGlys----- 3529
QY 8193 ACATTAAGTGGACAT-----AAGACGATCCACCTGTTGATGA 8231
Db 3530 ----GlnLeuGlnHisIleGlyAlaArgAsnIleLysAspAspProLysMet----- 3546
QY 8232 TTAATACCTCAACGTTTCGATGATGACGTGATGCTGATTAATTTGTGAAGCACT 8291
Db 3547 -----MetTrpSerIleHisValAlaLys----- 3554
QY 8292 GATGAGAAATGCAAAATTTAAATATCATGTGATCAGTGAATACATCGACAGATG 8351
Db 3555 -lIleGlnSerAspArgGlnTyLysGluPheGlnLysThr-----LysPh 3572
QY 8352 CAAGAATGATTTATGATGAATTAAGTGAACAGTGAACAGTGAACAGTGAACAGTGAAC 8403
Db 3572 eSerSerProValAspMetLeuGlnValValLeuAlaLys-----LysCysGlnIleLeuVa 3591
QY 8404 -----GAAATTAATAATTTGTTCTTAATGCAAACTCATATGAT----- 8445
Db 3591 lSerAspLysArgTyLysHisProLeuHisGlnIleTrpThrCysLeuProAspGlnAsnAs 3611
QY 8446 ----ATACATCAATTAATACAAAGAAATGTATGACACCAATATATACAAATCTC 8501
Db 3611 pValIleGlnAlaArgLysAlaTyArgLysLeuGlnSerAspAlaIleTyLys-----Se 3629
QY 8502 TACTATGATCATGTTCAAAATTT-----GTACAAA 8534
Db 3629 rAspLeuGlnIleTrpLeuArgGlyIleGlyTrpValProIleGlySerValGluValGlu 3649
QY 8535 GTTGAACCTTTAAAGTAGATGTCTGTGAGAGCTTTCTGAA----- 8580
Db 3649 sValLysArgAlaGlnGluIleLeuSerAspArgLysTyArgGlnProAlaSerGlnLe 3669
QY 8581 -----TATCTCATGCAACAGATAGTGTGTAATTAATTAATTAATGA 8630
Db 3669 uLysPheThrCysIleThrAspThrProGluIleValLeuAlaLysAsnAlaLeuTh 3689
QY 8631 TGGTTCCTCCAAATATGACACATATGCTTCGAA----- 8664
Db 3689 rMetSerLysHisLeuTyThrGluAlaIleTrpAspAlaAspLysThrSerIleHisValMe 3709
QY 8665 -----GAAACACCA-----AAAAGTTATTAAGAACTTGCAGTTGTACACT 8705
Db 3709 tProAspThrProAspIleLeuAlaLysSerAsnSerAlaAsnIleSerGlnLysSte 3729
QY 8706 ACCTTGAAGATCATGATGATATGCTACCGACGCAAAACAAAGATGATGAAGCA 8765
Db 3729 uTyTrThLysGly-----TrpAspGlnSerLys----- 3738
QY 8766 ATTAACAACCTTTACCTTCTGCTGGAAGATGATTAATGATTAATCTGATTAATGGA 8825
Db 3739 -----MetLysAspTyTrAsPheArgAlaAsp----- 3747
QY 8826 CGCATACCTGTTCTTAATAGTCAAGTGAATGATTAACAAAGGTGTATGATTCCTCAAGAG 8885
Db 3748 ----AlaIleSerIleLysSerAlaLysAlaSerArgAspIleAlaSerAspTyTrLys 3766
QY 8886 AAGACATTTATGTACAAGACCTATCACTGATTAATTAATGAAGAAAGCT----- 8934

[illegible]

Db	4046	gargleuhsprtrialals	-----	4052
Qy	9903	gtrbagsttgatnagsgcsttcattcatgacagaaaaattcaatgcstgsgacttctt	-----	9962
Db	4053	-----	-----	4064
Qy	9963	gcgtatpactgaattatcccgacagagacatggaatgacctagcttgaaatccaaaaatag	-----	100222
Db	4064	u8erlleuhts	-----	4072
Qy	10023	gtacatpaccatpatagaactgctccatpataaggcaaaaacatatatatat	-----	100711
Db	4072	rglnalaieuvalserasmetasprtyr	-----	4090
Qy	10072	-atggaagagactactagcgagatgacgaatgaatpataatgpcggacttatcttcctgtca	-----	10130
Db	4090	smectroasprglnasnarvalillelnalalsysalatytgltleuclnserasprss	-----	4110
Qy	10131	tattacttcattccgaagtgactat	-----	10163
Db	4110	nvaltyrtylsalaa8prleugltutprleuarglylleglytrmetproasnasprer	-----	4130
Qy	10164	ggatpattat	-----	10205
Db	4130	lseralalshnlisalaalynlshlalaalaspillepe	-----	4148
Qy	10206	aacattgatatgaactagactagacacacaaagagatataccaaactgattatgacac	-----	10265
Db	4148	ghrlrystllegltutprleu	-----	4163
Qy	10266	aaatgattgattacacacagct	-----	10307
Db	4163	lasprtyvalthrallalalsglnserglutlleuaspasprlileystyrarglysas	-----	4183
Qy	10308	atggaatgaaactgaactgaatttt	-----	10332
Db	4183	ptrapsnalatprlyserlystythrleuthrgltutprleuclnshlthralagl	-----	4203
Qy	10333	-----	-----	10376
Db	4203	ngluhlaalaargtllleuasp8lntutleu	-----	4215
Qy	10377	aaagtcagatattccaaatgaatnacsaaacsttaattcttatattctsgatnactcgaaaga	-----	10436
Db	4215	soluclly	-----	4232
Qy	10437	aaaacsttttatattatattcatattcatgattggaatttatattactggaaagaaattagta	-----	10486
Db	4232	aalrprophelvalnshlanlnshlscysahmsprvalglnser	-----	4250
Qy	10497	t 10497	-----	4250
Db	4250	t 4250	-----	4250

QY 1864 ATTGATGAATGTAACAGAAATTCCTTAGTCTTTTGACAGATGCGTTAAACAAAAGAGACA 1922
::: ::::: ||| |||||
Db 105 PhgIumEcIySerLysAsnIlethCysPheTyProIleValGlyLysGluArg 124
QY 1924 GAATGAATAGT-----ATAAAGAACTGTTCCAAAAAAGATATACAGCA 1974
LysThrLeuAspThrIleIleIleLysAsnValThrAsnAspHisValIleSer 144
Db 125 LysThrLeuAspThrIleIleIleLysAsnValThrAsnAspHisValIleSer 144
QY 1975 TCGATTATAGTAATATATATATATCTTTTGAAGTTATTTTAACTATATGATTA 203
::: ||||| :::
Db 145 AspMetHisSerAsnVal----- 150
QY 2035 CTTCACAAAGATGACAGCAAAATGGAAGAACTTATGCAAAATATA---AAAAGAAAAA 2091
::: || ||||| ::::: |||
Db 151 -----GlnGluLysAsnMetIleuIleArgAsnIleAspLysGluAsnLys 166
QY 2092 AATGAGTTTCCAAATTGCGAAAT-----AATRGAGCATTTTACAGAAATCAATACAA 2145
||||| ::::: ||||| |||||
Db 167 AsnAspIleGlnAsnValGlnGluLysIleGlnArgAspThrTyGlnLys----- 184
QY 2146 CTCCTGTAGATCACTTAAAGAAATCCACAGATA---TGTAAAGCATATATACAAAC 2202
||||| ::::: ||| |||||
Db 185 -----AspTyGlnSerAspThrLeuIleGluThrPheAspAsnThrAsn 201
QY 2203 GAAGATGTGAACATCCCATTAATGCAACAACAAACCCGTGTGTAAACCTGCGAAGC 2262
||| ||| ::::: ||| |||
Db 202 Gln-----GluAsnPhe-LeuLeuThrPheLeuLysArgCysLeuMetLysIlePheSe 219
QY 2263 ACCGAACCCACTAAAAATATATAAAGAAATGACACATACTTAAAGAGCGATACACAG 2322
|||
Db 219 rserProLysArgLys-----LysThrValValGlnLys 230
QY 2323 GAAGACAGAAATCGTGGTCTTCATTAATTGAAGAAAGACACAGAGATATATATA 2382
||||| ::::: |||||
Db 230 sLysHisLysSerAsnPheIleAsnSerIleuLysTyIleTyMetLysLeuTh 250
QY 2383 CGTGGGGGTAGAGAAAGCACTTCAAGGACAAATTTATGAGAAATATATATAACATCT 2442
250 r-----ProserAspSer--PheAsnLeuValArg-----Arg 260
Db 2443 AATCGTAATCTTGGTTTTCAAAT---GGACCATGTGATGCAAGACAGATGATGT 2499
::: ||||| ||| |||
Db 261 AsnArgAsnLeuAspGlnGluAspMetSerProArgAsp----- 273
QY 2500 ATACAAACAGATTTGTCGTAGAGAACTGAAATGGAAAGTGCATCCGAGACATGCGTAAA 2559
274 -----AsnPheValIleAspArgGlnGluGlnGluGlnGluGlnGluGln 290
QY 2560 GATCAGCAAGATGTTATTTATGCTCTCTAGAAAGACACATATATGTACATCCAAATTGAA 2619
::: |||||
Db 291 GlnGluGlnGlu-----GlnGluGlnGlu 298
QY 2620 CATTTACAAAGCATGATCATCCCATTAATGCTAATATTGTTGATGATTAGTAAATAT 2679
::: ::::: |||||
Db 299 GlnGluGlnGluGlnGluGluTy-----AspAspTyVal----- 309
QY 2680 TCCTTTGGGGGATGTCTTCTATACGCAAAATATGAAGCAACAGATATATCGAATG 2739
309 ----- 309
QY 2740 TATTAAGAAAGATTAACCTTAAAGGGCCCAAGAAAGTAACTGACCCAAAACACAGACA 2799
||||| ::::: ||| |||||
Db 310 TyIleGlnLysSerGlyAsp-----GluThrGlnGlnGlnGlnGlnGlnGlnGln 327
QY 2800 ACATATC---TGTCGACCTATACGTTTACAGTTTTCGAGATATAGT---GATATAATTCCA 2853
::: ::::: ||||| |||
Db 328 GluValGlnLysIleAsnSerSerGlnGluSerPheAsnAspGlnAspSerValGln 347
QY 2854 GGAAGATCTCTGGGAAAGAAACGATGACATGTATAAAGTCAAGACATTTGCGAAAT 2913
||||| ||||| ::::: |||
Db 348 AAlArgAsp-----GlyAspMetIleArgValAspGlnLysTyTyGlnLys 362

Db 914 --- AsnAsnLysGluHisAspTyrThrCysAspPheThrAspLysLeuAspLysThrVal 932
QY 4876 AACATTTGAGTAAGTAAGT-----TGCAAGGATTAAGTAAGGCGCAAC 4923
Db 933 ProSerThrIleAsnGlyLysLeuPheIleCysAspGlyHisLeuLysGluPheAsp 952
QY 4924 ACAGGTCTTTAAAGTGGCAAAACAAAGGCTTAATACATACATTAATGAAAGATG 4983
Db 953 ThrPheThrLeuLysCys---AsnValAsnLysThrGlnTyrProAsnIleGlu----- 969
QY 4984 ACAGTAAGATGCTTTTCTCCGTCGACATACGTAATGTTTTCATGCTATGGATGCC 5043
Db 970 -----IlePheProLys----- 973
QY 5044 AATTATACAGATCCAGAGTTAAAGATGAATAAGTGGTGGCAAAAGATGATGAAGTG 5103
Db 974 -----ThrLeuLysAspLysGluValLeuLysLeuAspLysIle 988
QY 5104 -----CGGCGCAACGAAAGGATCAAT 5124
Db 989 GlnTyrGlnMetPheSerLysPhePheLysPheAsnThrGlnAsnIleLysTyrLeuAsn 1008
QY 5125 TTGGGTCAATACTACAA-----GAAAAAAGAAAAAGAAAA 5163
Db 1009 LeuTyrProTyrTyrLeuIlePheProPheAsnHisIleGlyLysGluLeuLysAsn 1028
QY 5164 ATAAAAACCGTGGATGCCCAATATTTTATAGGTCGCCCTTGATGCTATGAA 5223
Db 1029 AsnProThrTyrLysAsnHisLys-----Asp 1037
QY 5224 TATAGTTTATGATTAAGATTAATCTAGTATGATTAATTTGGAGAT----- 5277
Db 1038 ValLysTyrPheGluIleAsnSerValIleSerProLeuSerSerIleAspLeuGly 1057
QY 5278 -----GAAAAACAAAGACCGAGAAATTTTGAAGAAATATTATTAAC 5319
Db 1058 LysLeuLeuAsnPheLeuAspThrGlnGluThrValCysLeuThrGluLysIleArgTyr 1077
QY 5320 AAAAAATGGAACATCACTTGGCAAGAAAGTACTAGTACTACAGAAATCCCGGTACT 5379
Db 1078 LeuAsnLeuSerIleAsnGluLeuGlySerAspAsnAsnThrPheSerValThrPheGln 1097
QY 5380 GCGCGAAATTTTCTGGAACGAAATTAAGATGCTGTGGAGCAATGATATGCGGG 5439
Db 1098 ValProProTyrIle-----AspIleLysGluProPheTyrPheMetPheGlyCysAsn 1115
QY 5440 TACAAACGTGTGATGATGAAATAGTGA-----AATAGTCAAGAAAGT 5487
Db 1116 AsnAsnLysGly-----GluGlyAsnIleGlyIleValGluLeuLeuIleSerLysGln 1133
QY 5488 GATGAAGATCTAAAAAATGCTTCTGACTTACATGATGATGATATTCCTATGGCGAA 5547
Db 1134 GluGluLysIleLysGlyCysAsnPheHisGlySerLysLeuAspLys----- 1149
QY 5548 AATCCGATGAAGTACTGCGTATCATGTTCTTCGATGGTTTGGCGAATGGGTGAAGAT 5607
Db 1149 ----- 1149
QY 5608 TTTTGCACAACATTAAGAAAGAAATTTGGAGAAATTTGGAGGCGCTGATCATATATCT 5667
Db 1150 -----PheAsnGluAsnIle 1154
QY 5668 TGTGCTGATTAATGAAGATAAAGAAATGTACAGATGCGGTGACACATATATAAATA 5727
Db 1155 SerSerAsp-----ThrHisGluCysThrLeuHis----- 1164
QY 5728 TTTATTTAGTGAAGTGAACCAAGTATGAATAAACCAATCAAAATATATGAGAGATTA 5787
Db 1165 -----AlaTyrGluAsnAspIleIleGlyPhe----- 1173
QY 5788 GACAAATATATTCGACATCT-----GTGGCAAAAGTGCAGAGACCGCTGCCGAA 5841
Db 1174 AsnCysLeuGluTyrThrHisProAsnGluValGluValGluValGluAspIleGlu 1193
QY 5842 TATTTGACAAACATTAATAAATTTGTGAATAAATGAAGATGATGATTAAG 5901
Db 1194 TyrLeuGlnProGlu-----AsnCysPheAsnAsn-----ValTyrLys 1206
QY 5902 TGTATGAAGATGTGTCACACAGCATTAATGATGATGAT-----AAT 5943
Db 1207 GlyLeuAsnSerValAspIleThrThrIleLeuLysAsnIleGlnThrTyrAsnIleAsn 1226
QY 5944 AGTCAAAATATGCCCGCATCATTAAGACATGAACCA-----AAAGAGTTGA 5991
Db 1227 AsnLysLysThrProThrPheLeuLysIleProProTyrAsnLeuGluAspValGlu 1246
QY 5992 GGAAGTGTAAATTTGCAATGCGCAGAGGTCCACACAGTGA----- 6033
Db 1247 IleSerCysGlnCysThrIleLysGlnValValLysLysIleLysValIleIleThrLys 1266
QY 6034 -----CGAAGGAAACACCGTCA----- 6051
Db 1267 AsnAspThrValLeuLeuLysArgGluValGlnSerGluSerThrLeuAspAspLysIle 1286
QY 6052 -----CCACGGTATCACTGATA----- 6069
Db 1287 TyrLysCysGluHisGluAsnPheIleAsnProArgValAsnLysThrPheAspGluAsn 1306
QY 6069 ----- 6069
Db 1307 ValGluTyrThrCysAsnIleLysIleGluAsnPhePheAsnTyrIleGlnIlePheCys 1336
QY 6070 ---TCAAAAGCAGCGCATCGAATAAGAAAGCAAA-----ACAGCGCGCT 6114
Db 1327 ProAlaLysAspLeuGlyIleTyrLysAsnIleGlnMetTyrTyrAspIleValLysPro 1346
QY 6115 ACAAAACGCGC-----AAAAAGTGAAGAT-----CTAACA 6147
Db 1347 ThrArgValProGlnPheLysLysPheAsnAsnGlnGluLeuHisLysLeuIleProAsn 1366
QY 6148 ACAGAAATGCGAGCAGCAACAGCAACCGAGCAGCAGCAGCAACAAACGAAAA----- 6201
Db 1367 SerGluMetLeuHisLysThrLysGluMetLeuIleLeuTyrAsnGluGluValAsp 1386
QY 6201 ----- 6201
Db 1387 LeuLeuHisPheTyrValPheLeuProIleTyrIleLysAspIleTyrGluPheAsnIle 1406
QY 6202 ---GCAACATCAACAGCAACAAAGAAATCTACCGT---GGCACAATGTGAAGGCC 6255
Db 1407 ValCysAspAsnSerLysThrMetTrpLysAsnGlnLeuGlyLysValIleTyrHis 1426
QY 6256 ATCTTTCGAATTAACAGATAGCAGGAGGTGAATAGAGGTGT----- 6300
Db 1427 IleThrValSerLysArgGluGlnLys-----ValLysGlyCysSerPheAspAsnGlu 1444
QY 6301 ---ATCCAAAAAGTATGAGCAATATCTTAATGGGCTGTATTTAGTAAGTCT 6354
Db 1445 HisAlaHisMetPheSerTyrAsnLysThrAsnValLysAsnCysIleIleAspAlaLys 1464
QY 6355 AAAGAAATGAATAATGCAATATGATGCTCTCTAGAGAAATAATTA-----TGN--- 6405
Db 1465 ProLysAspLeuIleGlyPheValCysProSerGlyThrLeuLysLeuThrAsnCysPhe 1484
QY 6406 -----ATAAATATATACAAATTTTA----- 6426
Db 1485 LysAspAlaIleValHisThrAsnLeuThrAsnIleAsnGlyIleLeuTyrLeuLysAsn 1504
QY 6427 -----AATTAAT---GAAATGAAAAATAAGCGT 6450
Db 1505 AsnLeuAlaAsnPheThrTyrLysHisGlnPheAsnAspIleTyrLeuProAlaLeuMet 1524
QY 6451 GACAATGATTAATAAGAGCTTTATTAATGTGCAGCAATAGAACTGAATTTTGTGG 6510
Db 1525 AspAsnAspIleSer-----PheLysCysIleCysValAspLeuLys-----Lys 1539

QY 6511 TTAATAATATATATGAAAAATCTCGCAGCAGAAAATGATTCGAA----- 6555
Db 1540 LysIleuTyraValIleuSerProLeuGlyProLysValIleuValIleuTyraLys 1559
QY 6556 -----AATGAAACAATTCAGATGATTAATTAAGAATA 6588
Db 1560 LeuAsnIleuLysPheAspAsnTyraIleuThrCylThr-----AspGlnAsnIleuTyraLys 1577
QY 6589 ATGTATATATACATATGCTGAT-----TATAAGATATG 6621
Db 1578 Met-----ThrTyraIleuAsnIleuSerHisLysAsnTyraLysGluLeu 1595
QY 6622 TTTTTCGAACGTATATTTTGAATAAAAAATTAACCTGACAAATAGCTGTACA 6681
Db 1596 PheHis-----AspLeuGlyLysLysProAlaSerPheAlaAsnProGluSer 1613
QY 6682 ACCATT-----CTCAATGAAATATATAGAAAAACGATTAATAAAAAAGAT 6729
Db 1614 IleIleGluSerLeuSerIleAsnGluSerAsnGluSerGlyProPheProThrGlyAsp 1633
QY 6730 GAAGAATTAACGTAATAATATTTTGGAGAAAAATTAATTTTGGAGAGATGATA 6789
Db 1634 ValAspIleGluHisLeu-----IleLeuGly----- 1643
QY 6790 TATGATTAATATATATCTCTCAGACAGAAAACGAAAAAATTAAGATATATAC 6849
Db 1644 TyraSerThrTrpIleuSerLeuTyraAspGluIleuGluValIle----- 1659
QY 6850 CACTACATGACATGACCAAACTG---ACGCCTCCCTGAGAGTTGTAAAAAGGCC 6906
Db 1660 ---TyraAsnIleuGluSerLeuGluLeuLysAspIleuGluIleuTyraValIleGluVal 1678
QY 6907 CAATTTTTCAGATGTTTACAGAAATGGCGAAGAAATTTGTAAATAGAGAGAACAG 6966
Db 1679 AsnLeu-----LysAlaProLysLeu 1685
QY 6967 TTGTTAAATTTGAGGGGGCTGTAAAGAAATAGCTGTAATGTAATGACGGTAAG 7026
Db 1686 MetIleSerAlaGlnIleHisAsnAsnAlaGlnIleValLysAspPheSerLysAsn----- 1703
QY 7027 ACACAGAAATGTCAGAGCGGTGTAAACATATCAAAATTTTAAAGATGGAAAACT 7086
Db 1704 -----AsnIleuIleValPro----- 1708
QY 7087 GAATATGAAAAGACAAAAGAAAGTTCAAAAAGGATTAAGTGGCAAAAAGTATAGGAT 7146
Db 1709 -----GluSerLeuLysLys----- 1714
QY 7147 TATCCTTCTACTGAAGACATTAAGACAGCAACATGCTCATGAATATTTAAACATG 7206
Db 1714 ----- 1714
QY 7207 AAATTAAGAATATATGTCGAATAG-----GATTGCTCTTATGCAAAAACCT 7257
Db 1715 -----GluIleuGluGlyLysAsnProValAsnIleHisCysTyraAlaLeuLysPro 1732
QY 7258 TCTTCACACATA-----CCAAAAACAACACAAATCACAATCATCCGATGCT 7305
Db 1733 LeuAspThrLeuTyraValLysCysProThrSerLysAsnTyraGluAlaIleVal 1752
QY 7306 AATGATATGCGACAAATGCGGATATGTCCTCGAACAATTTAAACAAGTGTGATGCT 7365
Db 1753 Asn-----IleSerGlnAsnAspAsnGluTyraGluLeuGln 1764
QY 7366 GAACCTTCAAAAAAGGATCATGATCATCAAAAAAATTAAGTGAACCTTAATACCT 7425
Db 1765 ValIleSerIleuIleGluLysArgPheHisAsnIleuThrIleuGluSerLysPro 1784
QY 7426 ATGATATGTGTAGAGAAAGACATATTTATCTTAAGACAGCAAAAAATATATGAT 7485
Db 1785 -----GlyAsnGlyAsp 1788
QY 7486 ATTACCTTGAAGAAATTTATA-----CCTATT-----GAGCTACAAAGGA 7530

Db 1789 ValValAlaHisAsnGlyValIleAspThrGlyProValIleuAspAsnSerThrPheGlu 1808
QY 7531 AAGGAAAGTAAATAAGTGTGACTAATTAATTAATCTTCGATCCTTAAGAAACCTTATGCA 7590
Db 1809 LysTyraPheLysAsnIle-----LysIleLys 1817
QY 7591 CTTGATTAATATATATAGAGACA---AGAAACCTTTCGAAATATAGAGAAATAATCGTTT 7647
Db 1818 ProAspLysPhePheGluLysValIleAsnGluTyraAspPheThrGluGluLys--- 1836
QY 7648 AAGGTAGATTATGAA-----TGGAAATGTTACAAA 7677
Db 1837 -----AspLeuGluSerIleLeuProGlyAlaIleValSerProMetLysValIleLys 1854
QY 7678 AATTCAAAAGTTCTATCAGAGAAAAAAGATATGTTGATCCCAAGACAGAACATATG 7737
Db 1855 LysLysAspProPheThrSerTyraAlaIlePheValIleProPro-----Ile 1870
QY 7738 TCGTTAAGAAATTTAGATGAATTAATAATGAA-----ACACTTAAGAGATAGT 7785
Db 1871 ValProLysAspLeu---HisPheLysValGluCysAsnAsnThrGluTyraAspGlu 1889
QY 7786 AATTATCTCTAAAAATGCTTCGTCGAACTGCAGCAATAGCAATAGCAATATATATAA 7845
Db 1890 AsnGlnTyraIleSerGlyTyraAsnGlyIleIleHis-----IleAspIleSerAsn 1906
QY 7846 AACTTCAACTCAGAGAGGGGTGGCAATGAAATCAATATGTGATTAATAATATAGT 7905
Db 1907 SerAsnArgLysIleAsnGlyCys-----Asp 1915
QY 7906 TTGCGTATGTCGGTGCATATAGTAGAGAACAGATATGTAATGTCGTGCTTACTTA 7965
Db 1916 PheSerThrAsnAsnSerSerIleLeuThrSerSerValIleuValAsnGlyGluThr 1935
QY 7966 CTTCCCGTAGAAATTAATAATTA----- 7986
Db 1936 LysAsnCysGluIleAsnIleAsnAsnAsnGluValPheCylIleIleCysAspAsnGlu 1955
QY 7987 -----TATAAGCTTTTGAATATATATATATGAAATAGAGAAATTAATAAT 8031
Db 1956 ThrAsnLeuAspProGluLysCysPheHisGluIleTyraSerLys-----AspAsn 1972
QY 8032 AAAGGTGAATTAATTAACAACGATGTA-----CAACGTTTCTGCTGCTGG 8079
Db 1973 LysThrValLysLysPheArgGluValIleProAsnIleAspIlePhe-----SerLeu 1990
QY 8080 TGGATGCTAATATGAAGAATATTTGGAAGCA----- 8112
Db 1991 HisAsnSerAsnLysLysValAlaTyraAlaLysValProLeuAspTyraIleAsnLys 2010
QY 8113 -----ATGACGTCCAAAGCACAGAAAGATGCAAAACCTTTTAAAGAAAGAGA 8160
Db 2011 LeuLeuPheSerCysSerCysLysThrSerHisThrAsnThrIle----- 2025
QY 8161 ATGAGATGATTTGAACCATATACATTAATACAGATATAGTGTGACATTAAGACGATGCA 8220
Db 2026 -----GlyThrMetLysValThrLeu-----AsnLysAspGluLys 2037
QY 8221 CTTGTTGATATATATATATACCTCAACGTTTCGATGATGACTGAATGTGATATATAT 8280
Db 2038 GluIleuLysPhe----- 2042
QY 8281 TGTAAAGCACTGATGAGAGAAATTTAAAAAAATCATGTGATCAC---TGTAAA 8337
Db 2043 -----LysThrAlaGlnGlyLysHisHisAsnValHisLeuCysAsn 2057
QY 8338 ACATCTGACAGATGCAAGATGATTAATGATGAGAAATATAGTGTGAACGCTGTA--- 8391
Db 2058 PhePheAspAsnProGluLeuThrPheAspAsnAsnLysIleValIleuCysLysIleAsp 2077
QY 8391 ----- 8391

```
Db 2078 AlaGluLeuPheSerGluValIleIleGlnLeuProIlePheGlyThrLysAsnValGlu 2097
Qy 8392 -----ACGAGATGTCAGAAATATATAAAATTTGTTCTTAATGGAATCTCATTCGAT 8445
Db 2098 GluGlyValGlnAsnGlnGluIleLysLysPheSerLeuLysProSerLeuValPheAsp 2117
Qy 8446 ATACAAATCAATTAATACAAAGAAATGTAATACACCAATATATATACAAAATCTCT-- 8502
Db 2118 AspAsnAsnAsnAspIleLysValIleGlyLysGlnLysAsnGlnValSerIleSerLeu 2137
Qy 8503 -----ACTTAAATGATGTTCAAAATTTGTACAAAAGTTGAAACATTTT-- 8547
Db 2138 AlaLeuLysGlyValTyrGlyAsn-----ArgIlePheThrPheAsp 2151
Qy 8548 AAAAGTAAAGTCTGTTGAGAGCTTTCTGAAATATCTTCATGAAACAAAGTATGTTTG 8607
Db 2152 LysAsnGlyLysLysGlyGlnGlyIleSerPhePheIleProIleLysGlnAspThr 2171
Qy 8608 AATATATAATTT-----AATGAAAATGATGTTCTTCCAAATATACGA----- 8649
Db 2172 AspleuLysPheIleIleAsnGlnThrIleAspAsnSerAsnIleLysGlnArgGlyLeu 2191
Qy 8650 ACATATGCTTTGCAAGAAACACCAAAAAGTTTAAAGAAAGCTTGACGTGTACATACCT 8709
Db 2192 IleTyrIlePheValArg----- 2197
Qy 8710 TCTAAGATTCATTCGATTAATTCCTCCACCAACAAAGATGAGTAAAGAAATTA 8769
Db 2198 ---LysAsnValSerGlnAsn----- 2203
Qy 8770 CAAACTTTTACCTCTGCTGCAAGAAATGATTAATTAATCTGATAATTTGAAACGCA 8829
Db 2204 ---SerPheLysLeuLysAspPheThrThrGlySerThrSerLeu----- 2217
Qy 8830 TACCTGTTCTTAATAGTTCACATGATTAACAAGGTGATTAATTCCTCCAAAGAGAGA 8889
Db 2218 ---MetGlnLeuAsnSerGlnValLysGlnLys----- 2227
Qy 8890 CATTTATGTAACAGACCTTCATGCTGATTAATTAAGAAAGGTATTAAGAAATTTTA 8949
Db 2228 ---LysCysThrValLysIle-----LysLysGlyAspIlePheGlyLeu 2241
Qy 8950 AAAAAAACTCTTACTTCTGCTTCAAGTCA-----GACAAATTTAGTGTCA 9000
Db 2242 LysCysProLysGlyPheAlaIlePheProGlnIleAsnLysPheSerAsnValLeuGlu 2261
Qy 9001 AATATATAATCGAAGAAAGTGTGCTTGAAGCAATTAATTAATGATTAATGCAATTA 9060
Db 2262 TyrTyrLysSerAspTyrGlnAspSerGlnHisIleAsnTyrTyrIleHisLysAspLys 2281
Qy 9061 TCCGATATATTAAGAAAGTAT-----ATGATGACACATTCATTAATCTGAA-- 9108
Db 2282 LysTyrAsnLeuLysProLysAspValIleGlnLeuMetAspGlnAsnPheArgGlnLeu 2301
Qy 9109 -----AAAAATTAATAATTTGAAACATCAAT 9138
Db 2302 GlnAsnIleGlnIleTyrThrGlyLysSerAsnIleThrAspValLeuHisPheLysAsn 2321
Qy 9139 GAAGCAACCGAATATCGTAACATGATGGGAAATTAATAGCCTCAATATGCAACCT 9198
Db 2322 PheAsnLeuGlyLysLeuProLeuAsnPheLysAsnHis----- 2334
Qy 9199 ATGTTATGTCATTAATAATTTGCTACTCAAAAGTACATTAAGTAAAGATGATGTCAA 9258
Db 2335 -----TyrSerThrAlaTyrAlaLysValPro----- 2343
Qy 9259 TTACCAAAAGATGAAGAAGTAATAGTTCTGCTGTTAATTAATGATGGCAAGCA 9318
Db 2344 -----AspThrPheAsnSerIleIleAsnPheSerCysAsn 2355
Qy 9319 GCATGTAAGAAAGAAACATGATAGTATCATTAATAAACAATATGCTCGTTCAAC 9378
Db 2356 CysTyrAsnProGlnLysHisValTyrGlyThrMetGlnValGlnSerAspAsnArgAsn 2375

Qy 9379 GAAGTAAATTTAGACGCTGAGATTAATTAAGACAACTGGATGTCAGATGATTAAG 9438
Db 2376 PheAsnHisThrTyrIleLeuLysAsnGln-----AsnValIleLys 2387
Qy 9439 AATATATTAATGCTTGAATATTAATGATAAAATATCAATGGAATAATCTAATTAATTA 9498
Db 2388 AsnPheLeuLeuProAsnIle-----GlnLysTyr 2397
Qy 9499 AACCAATTAAGAT-----CAATCTGAGTAAATATAGACAAATTAACATCTGAGAA 9552
Db 2398 AlaLeuLeuLeuAspArgGlnIleArgGlnLysIleLysGlnGlnGlnGlnGln 2417
Qy 9553 AATGTCAGTCAATATTAATAACAAAGATTTGCAATGCGCTTGGAG----- 9600
Db 2418 GlnGlnGlnGlnIleLeuLysAspGlnAspArgGlnSerArgHisAspArgAsn 2437
Qy 9601 -----TTAAATGATTAATTAAGAAATGTAACAGAAACAAATAAT 9642
Db 2438 LysAsnHisThrTyrIleLeuLysAspSerAsnGlnHisIleCysAspTyrGlnLysAsn 2457
Qy 9643 GAAAT-----AATGAATTCAAAGATTAATTAATAATTAATATCT 9684
Db 2458 GluSerLeuIleSerThrLeuProAsnAspThrLysIleGlnLysSerIleCys-- 2476
Qy 9685 GGTATATATTTGTGAAGATGAACACAAACAAATATCATGATGATGAAATTAATA 9744
Db 2476 ----- 2476
Qy 9745 GAAGAAAGCAAAACAGTTCGCTTAAGCACTCATTTCTTTACA-----CCCAT 9795
Db 2477 -----LysIleAsnAlaLysAlaLeuAspValValThrIleLysCysProHis 2492
Qy 9796 GTAGATCTTTCTATCAACAGCACTTATTCACACATCAGTACAGTAAATATGATCCT 9855
Db 2493 ThrLysAsnPhe-----ThrProLysAspTyrPhePro 2503
Qy 9856 AAAAAATGATTAATTAAGTATGATCTGTTGATTAATGATGCGGTAGTTGATA 9915
Db 2504 AsnSerSerLeuIleThrAsnAspLysLysIleValIleThr----- 2517
Qy 9916 GCGCTTATTCATGAGAAAGAAATTAATCAATGCTGTGAGCTGTTGCTATGATGAT 9975
Db 2518 -----PheAspLysLysAsnPheValThrTyrIleAspProThrLys----- 2531
Qy 9976 ATCCGCAAGAGAGATGGAATGCTGATGCAATCCAAATAATAGTACATACATAT 10035
Db 2532 -----LysThrPheSerLeuLysAspIleTyrIle----- 2541
Qy 10036 AGAAGTGCATATTAAGCAAAACATATATATATATGGAAGAGATAGTATGAGAT 10095
Db 2541 ----- 2541
Qy 10096 GAAGTAAATATATGTTGGAATTAATCTTCTGATATTAATCTCATCCGAATGAGTAT 10155
Db 2542 -----GlnSerPheTyrGlyValSerLeuAspHisIleAsnGlnIleLysIleHis 2559
Qy 10156 GAAGAAATGATTAATTAATGATATATATATGATACCAAGTATGCTTAATTAATAACATGATA 10215
Db 2560 GlnGlnIleTyrAsp-----AspValHisLeu-----PheTyrProProHisAsnValLeuHis 2576
Qy 10216 GAAGTACTACTAGAACCA----- 10233
Db 2577 AsnValValLeuAsnAsnHisIleValAsnLeuSerSerAlaLeuGlnGlyValLeuPhe 2596
Qy 10234 TCAAAAGGATATACCAAGTATGATGATACCAAGTAAATGATACACACGCTACGATTAAGA 10293
Db 2597 MetLysSerLysValThrGlyAspGlnThrAlaThrLysLysAsn----- 2611
Qy 10294 TTATATGATCATGAAGAATGAAGTAAACATGATTTGTATCTCAAAATTAATTAACCAAT 10353
Db 2612 -----ThrThrLeuProThrAspGlyValSerSerIleLeu----- 2623
```


Protein

2107 -----TTGAAATAATTAAGGACTATTACG-----AATGCATAAGAATCTTGG 2151
Db 91 GlnValylserLeuSerGlnValAsnPhgInglngLysIleuAsnSerGlyLysGln 110
QY 2152 TTAGATCATTAAGAAAGAAAGTCGCCAGATGTGAAGACAANT---AATACAAACGAGCA 2208
Db 111 ILeuLysLeuGlnGlnGlnIleuLysArgCysLysSerGluLeuGluArgSerGln 130
QY 2209 TGTGAATCATCCCATATATGACACAAACCCTGTGTTAACTCGTGGA---GGCACG 2265
Db 131 AlaAlaGlnSerAlaAspValSerLeuAsnProCysAsnThrProLysIlePheThr 150
QY 2266 CAACCCACTAAAAATATAAAAGAAATAGACACATCTTTAAAAAGAGTCGATCGAGGAA 2325
Db 151 ThrProLeuThrPro-----SerGlnTyTrSerGlySerLysTyrgluAsp 166
QY 2326 GCACGAAATGCGTGCTTCATATATGAAAGAAAGCACACAGAGATATATATAAGT 2385
Db 167 -----LeuLysGluLysTyrsAsnLysIleValAlaGlnGlu--- 177
QY 2386 GGGGGTAGGAAAGACATTCAGACAAATTTATGTAGATATATGATAAACATCTTAT 2445
Db 178 -----ArgLysArgLeuGlnAlaGluValLysAlaLeuGlnAlaLysLysAlaSer 194
QY 2446 CGTAACTTGCTTTTCAAATGACCATGTGATGCGAAAGCAGCGTGATGTATACAA 2505
Db 195 GluThrLeu----- 197
QY 2506 ACAGATTTGTGTAGAGACTGAATGGAGATGCGAGACACATGCGTAAAGTAC 2565
Db 198 -----ProGlnAlaThrMetAsnHis 204
QY 2566 GAAGATGTTATATGCTCTCTAGAAAGACATATATGTACATCCAATTTGGACATTTA 2625
Db 205 ArgAspIle-----AlaArgHisGlnAlaSerSerSerValPheSerTrp 219
QY 2626 CAAAGGATGATCACCCA-----CTTAATGATATATTTGTTGATTTAGTATATAT 2679
Db 220 GlnGlnGlnLysThrProSerHisLeuSerSerAsnSerGlnArgThrProLysArg 239
QY 2680 TCCTTTTGGGGGATGTTCTTCTATCAGCAAAATATGAGCAAAACMAATATACGATG 2739
Db 240 AspPhe-----SerAlaSerTy 245
QY 2740 TATAAAGAAAGAAATACCTAAAGGCCCCAAAGAGATGACGCCAAACACACAGCA 2799
Db 246 -----PheSerGlyLeuLeuGlnValThrProSerArgSerThrLeu 259
QY 2800 ACTATC---GTTCCAGCTATACGTTACGTTTT-----GCAGATATA 2838
Db 260 GlnIleGlyLysArgAspAlaAsnSerSerPhePheGlyAsnSerSerProHisLeu 279
QY 2839 GGTGATTAATTCGAGAAAGATCTCTGGGAAGAAAC-----GGTGACATG---GTA 2889
Db 280 LeuAspGlnLeuLysAlaGlnAsnGlnIleuLysArgAsnLysIleAsnGlnLeuGlnLeu 299
QY 2890 AAGTCGAAGACATTTTGGAAACTGTTTTGGTATATACATTAAGTACTCAACAGGCAAA 2949
Db 300 ArgLeuGlnGlnHis-----GluLysGluMetLysGlyGln 311
QY 2950 GGAATGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3009
Db 312 ValAsn-----LysPheGlnGlnLeu 320
QY 3010 GAACCTAATAGCCAAAGATGAGGAGCCATG-----AATGCT 3048
Db 321 GlnLeuGlnLysAlaLysVal---GluLeuIleGlnLysGlnValLeuAsnLysCys 339
QY 3049 GATTTAAATATTTGAAGGATTAATCGGACACCAATCAACACAAAGTACTGTTGGGA 3108
Db 340 ArgAspGlnLeuValArgThrThrAlaGlnTyAspGlnAlaSerThrLysTyThrAla 359
QY 3109 TATAGTATATACACCATGATGATGATATATCCCAAAATAATTAAGATGATGACCGAA 3168

Protein

360 -----LeuGlnGlnLysLeuLysLysLeuThrGlu 366
QY 3169 TGGCAGATGTTACTGCAAGTCGAGAAAAAGAG-----TATGAT 321
Db 370 -----AspLeuSerCysGlnArgGlnAsnAlaGlnSerAlaArgCysSerLeuGln 387
QY 3211 AAGTTGAAGAGAGAGTGTAAAGAGATGAAGATTAAGTATGTCAGAGCTGACAAA 3272
Db 388 LysIleLysGlnLysGlnLysGlnPheGlnIleu----- 398
QY 3271 GAGAGTGTACAGGTTGTACAGAAAGTCACAGACCTTGTAATGAATTAATGAATGA 3333
Db 399 -----GluLeuSerArgGlnGlnArgSerPheGln 408
QY 3331 GGATTTGAAGAAACAATGGAATATATATACATTAATTAACAAAGATTAATACATGA 3399
Db 409 ThrLeuAspGlnGlnLysIleGlnMetLysAlaArgLeuThrGlnGlnLeuGlnAla 428
QY 3331 GCACAAATG---TGTGTTAGTAATAGTGTATTAAGCTTCAGATGCGCAAAATCAT 3444
Db 429 LysAsnMetHisAsnValLeuGlnAlaGlnIleuAsnProLysLeuThrSerValLysGln 448
QY 3448 ATAGCAGGAATGTTATGATTTTTCGGAATTAATCAACAAATGTCGCCAAAAGT 3500
Db 449 LeuGlnAsnLeuGlnGlnLysGlnLysLeu----- 460
QY 3508 AATAAAGTGTACTAGTATGAAGAAAGTCTGTCTATTGTTACTAACACACAGTATGA 356
Db 460 ----- 460
QY 3568 GTTGACACATATCTCATGATACAGAAATTTGATGATGTCAGTCACAAAATGAGTT 3622
Db 461 -----AspLeuSerGlnGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 467
QY 3628 TGATGTAAGAAAGTGTAGTGAAGATTAACGAAATATATGCTTTAGATTAACACAG 368
Db 468 -----GlnAlaSerGlnIleLysGlnAlaGlnIleu----- 476
QY 3688 GAGCATGATGTCGCTGTGTTAAAGTGGATGCAAAACCGCAAGGTTACAGATMAA 3744
Db 476 ----- 476
QY 3748 AGCAAAAAAAAAAGCGAGAAAGAAAGATGAGATGAATAACAGTGAATGATTAAT 3800
Db 477 LeuArgArgSerMetGlnGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 492
QY 3808 GAAAGATGGAAGAAACAAGATGACATGTCATCCAAA-----AAGATAGT 3856
Db 493 SerHisSerGlnIleLysAlaArgValCysHisLeuGlnAlaGlnLeuLysAsnIle 512
QY 3859 MATGATATCCGATGCGATGCGCAAT---ATTAATTTAGTGAAGACCCCTG 3912
Db 513 Lys-----GlnLysLeuAsnGlnSerGlnAsnPheAlaGlnIleuLys 527
QY 3913 CTGTGATGCCCCCTAGAAAGCAAAAGTTATGCGTATCTTGGCAATGATGA 3972
Db 528 -----AlaLysAsnThrSerGlnGlnThrMet 536
QY 3973 ATTAATAATTAATCAATCAATTAAT-----TTA 4002
Db 537 LeuArgAspLeuGlnGlnLysIleAsnGlnGlnLysAsnSerLeuThrLeuGlnLysLeu 556
QY 4003 AAAGAGCTTTCATCAAAATCTGACAGACAGAAACATCTTCTCATGATTAATTAATA 4062
Db 557 LysLeuAlaVal-AlaAspLeuGlnLysGlnArgAspCysSerGlnAspLeuLysLys 576
QY 4063 A---GTAAGAGTGGGAAGGAATTAATCAATTAATTAATTAATTAATTAATTAATTA 4119
Db 576 ArgGlnLysHisIleGlnGlnLeuAsn-AspLysLeuSerLysThrGlnLysGlnSer 595
QY 4120 CCCGATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4170

D	b	596	LytlalLeuLeuSerAlaLeuGluLeuLysLysGluLysGluLeuLysGluGlu	615
O	y	4171	-----TTATTTGGACAGACATATATGCAAAAGCTCATGTGTGAGGAGTAACATAAAAGAG	4224
D	b	616	LysThrLeuPhe-----SerCysTrpLysSerGluAsnGluLysLeuThr	631
O	y	4225	CAATATGATGTTCTCTTTCAAAATGGTGAGCAACCAAAATATCTCCATATGCGAAACACGCCCA	4288
D	b	632	GlnMetGluSerGluLysGluAsnLeuGlnSerLysIleAsnHisLeuGluLysLeu	651
O	y	4285	GAAATGTGACAGACAAACATAGCTATGAGATGTGGAGAGCTATGCTATGCTACTATATAAA	4344
D	b	652	LysThrGlnGlnIleLysSerHisGluTyAsnGluAlaArgThrLeuGluMetLys	671
O	y	4345	ATTGGGCGCAAAAAGATGATTTTACCGAAACAACTACGCTTACACAAACGTCAAAATTTAGT	4404
D	b	672	-----ArgGlnAsnLeuSerValGluIleArgAsnLeuHisAsnVal---LeuLys	687
O	y	4405	GACAAAGCACACACTTTGGAGGAAATTTGCCAAACGACCCACAGTTTATAGATGCGCTAAC	4466
D	b	688	SerLysSerValGluValGluThrGlnLysLeuAlaTyrMetGluLeuGlnGlnLysAla	707
O	y	4465	GAAATGTACGACGACATTTGCTATATACAGCAGCAATATTTGAGAGTATGCGACGAGAAA	4524
D	b	708	GluPheSerLys-----GlnLysHisGlnLysGluLysGluLysMet	721
O	y	4525	TGTAAAGTCAAAATGACCAATGTAAGGTGTGATACAGAAATGTAATTAAGAAATGCGAGACTAC	4584
D	b	722	CysLeuLysThrSerGlnLeu-----ThrGlyGlnValGluLysPhe	735
O	y	4585	GTT---AAATATATGAAAAAAGATGAGTGTCCCAAGATTAATATTTACAAGAT	4644
D	b	736	GluHisLysLeuGlnLeuLeuSerAsnGluIleMetAspLysAspArgCysTyrGlnAsp	755
O	y	4641	-----	4644
D	b	756	LeuHisAlaGluTyrGluSerLeuAlaGAspLeuLeuLysSerLysAspAlaSerLeuVal	775
O	y	4642	-----GAACGCGACAAAAAAGATTCGATAGACA-----CAC	4674
D	b	776	ThrAsnGlnAspHisGlnArgSerLeuLeuAlaPheAspGlnGlnProAlaMetHisHis	795
O	y	4675	ATTGGTGAATGCTTTACAGACCTTACTGCGACCAATATGCAACAGATTACTTGAACAGAAA	4733
D	b	796	SerPheAlaAsnIleIleGluGlnGlnGlySerMetProSerGluArgSerGluCysArg	815
O	y	4735	TTTACTGCTAGTGTGTGTGATTAAGCGGAGGCGCTCGTGTGTAACAAAGAAATTTACAA	4794
D	b	816	LeuGluAlaAsp-----GlnSerProLysAsnSerAlaIleLeuGlnAsnArgValAsp	833
O	y	4795	TTTGTAGAA-----AAACAGCTTACTATGATCCGCACAAACATGTT	4836
D	b	834	SerLeuGlnPheSerLeuGlnSerGlnLysGlnMetAsnSerAspLeuGlnLysGlnCys	853
O	y	4836	-----	4833
D	b	854	GlnGluLeuValGlnIleLysGluGluIleGluGlnAsnLeuMetLysAlaGluGlnMet	873
O	y	4837	-----CGGTGCAAAAATTTTGAATAATGACGAC	4866
D	b	874	HisGlnSerPheValAlaGluThrSerGlnArgIleSerLysLeuGlnGlnLysPheSer	893
O	y	4867	AAATATATCAATCATTT-----TCGAGTAAAGATTAAGTGCAAA	4902
D	b	894	AlaHisGlnAsnValAlaGluThrLeuSerAlaLeuGlnAsnLysGluLysGluLeu	913
O	y	4903	CGATTTAGTAAAGAG---GCAAAACACAGTGTATTAACTGCGCAAAACAAAGCTCTAAT	4955
D	b	914	GlnLeuLeuAsnAspLysValGluThrGlnGlnAlaGluIleGlnGlnLeuLysLysSer	933
O	y	4960	AACATAC-----AATTAAGTGAAGAA-----TTGACATGAAGATGCTTTTCT	5004
D	b	934	AsnHisLysLeuGlnAspSerLeuLysGluLeuGlnLeuLeuSerGluLysThrLeuSerLeu	953

Oy	5005	TCCTGCGACAGCTAATATCTTTTCATGCATTCGGAATTAATACAGATCCGAAGTT	5064
Db	954	GlutylusglutetserSerIleIleSerLeuAsnLysArgGluIleGlu---GluLeu	972
Oy	5065	AAAGATAAATGGCTGGCCAAAAGATTCATGGAAGTCGGCCGACCGAAGCGGTACAT	5124
Db	973	ThrGlnLysnGlnYthrIleuLysGluIleAsnAlaSerLeuAsnGlnGluLysMetAsn	992
Oy	5125	TTGGGCTAA-----TCTCAAAAGAAAAAGAAAAAGAGAAATA	5166
Db	993	LeuIleGlnLysSerGlnSerGlnSerPheAlaAsnTYrIleAspGluArgGluLysSerIleSer	1012
Oy	5167	AAACGTGCGATCGCGACAAATATCTTATGAGGTC-----CCGCTTGTAGTCATG	5220
Db	1013	GluLeuSerAspGlnTYrLysGlnGluLysLeuIleLeuLeuGlnArgCysGluGlnIthr	1032
Oy	5221	AAATATAGTTTTATGATTTTAAGAT-----	5247
Db	1033	GlyAsnAlaTYrGluLysSerGlnLysTYrLysAlaAlaGlnGluLysAsnSerLys	1052
Oy	5248	-----ATAATTCAGGTATTCATATTTGGAAAGATGAAAAACAAAGACCGAGAA	5298
Db	1053	LeuGluLysLeuLeuAsnGluCysThrSerLeuCysGlnAsnArgLysAsnGluLeuGlu	1072
Oy	5299	AATTTGAAGAAATATTTAACAAAAATGGAACATCAGTGGCAAGAAAGTATGACT	5358
Db	1073	GlnLeuLysGlnAlaPheAlaLysGlnLHisGlnGluPhe-----	1085
Oy	5359	ACAGGAAATCCCGGTACTACTGCCGCCAAATTTTTTGTGACCAAAATATAC-----	5409
Db	1086	-----LeuThrLysLeuAlaPheAlaGluGluArgAsnGlnAsnLeuMet	1100
Oy	5410	-----GATGTGTGTGAACGCAATGATCGCGGTACAAACGTCGTAGGATGAT	5460
Db	1101	LeuGluLeuGlnIthrValGlnGlnAlaLeu-----ArgSerGluMetThr	1115
Oy	5461	CGAAATACGTGAATTAAGTCCACAGACGTGAAGATCTAAAAAATTTGGTCTGTACT	5520
Db	1116	AspAsnGlnAsnAsnSerLysSerGlnAlaGlyLeuLysGlnGluIleMetThrLeu	1135
Oy	5521	TCACATGATGATTCATCTTATGGGGGAAAAATCCGAGATGACAGTCGATCAGTCTT	5580
Db	1136	LysGluGluGlnAsnLysMetGlnLysGluValAsnAsp-----	1148
Oy	5581	CGATGTTTCCGCAATGGGTGAAGATTTTGCAA-----CATAAAGAAAGCAATG	5634
Db	1149	-----LeuLeuGlnGlnAsnGlnLeuMetLysValMetLysThrLysHisGluLys	1166
Oy	5635	GAGCAATTCGAGGGCGGTATTCATTCATCTTGTCGATTAATGAAGATTAAGAAG	5694
Db	1167	GlnAsnLeuGlnSerGlnProIleArgAsnSerValLysGluArgGlnSerGluIthrAsn	1186
Oy	5695	AAATGTCACAGTCGCTACACAATTAATAAATTTATTAGAGATGGAACACACAGAT	5754
Db	1187	GlnCys-----AspPheLysProGlnMet	1194
Oy	5755	GAAAAACAATCAAAAATATGTTGTAAGAAATTAAGACAAATATATTCGCAATCTGTG	5814
Db	1195	AspLeuGlnValLysGlu-----Ile	1201
Oy	5815	GCAAAAGTCAGACGACGCTCGCAATTTTAGACAAACATTAATAAATAATTTGTGAA	5874
Db	1202	SerLeuAspSerTYrAsnAlaGlnLeuVal-----GlnLeuGlnAlaMetLeuArg	1218
Oy	5875	AATAAA-----ACTGAGATTTGGAATATTAAGTCTATGAA-----	5910
Db	1219	AsnLysGluLeuLysLeuGlnGlnSerGlnLysGluLysGluCysLeuGlnHisGluLeu	1238
Oy	5911	-----GATGTGTCACACACAGCATTAACGTGATGGTAATATGTCAAATATG	5955
Db	1239	GlnIthrIleArgGlyAspLeuGlnIthrSerAsnLeuGlnAspMetClnSerCln-----	1266

QY 9643 -----GAAATTAATGAATCAAGAGTA-----CTAAAAAA 9675
 Db 2559 AlacysLysalaIalysgluInaSnLeuSerGluValaIuIcysLeuGluValys 2578
 QY 9676 -----TTATATCCCTGCTTATATTTGTGAGATGAAGAACACAAATATCTATGTA--- 9726
 Db 2579 AlagInLeuLeuGluInglYleu-----AspIuValaIalysaSnasnyIYlleVal 2594
 QY 9727 -----CTAGATGGAATATATAAAGAGAG 9753
 Db 2595 LeuGInSerSerValaSnGluInglYleuGluValaIuIcIu 2607
 RESULT 6
 PVDB.PLAKN STANDARD: PRT: 1153 AA.
 AC P50493;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Duffy receptor, beta form precursor (Erythrocyte binding protein).
 Plasmodium knowlesi.
 Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 NCBI_TaxID=5850;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92357776; PubMed=1496004;
 RA Adams J.H., Sim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.;
 RT "A family of erythrocyte binding proteins of malaria parasites."
 RL Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).
 CC -1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
 DETERMINANT.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: HIGH TO P.VIVAX DUFFY RECEPTOR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: M90694; AAA29603.1;
 KM Multigene family.
 KN Multigene family.
 FT SIGNAL 1 21
 FT DOMAIN 22 1153
 FT TRANSMEM 1086 1106
 FT DOMAIN 1107 1153
 FT CARBOHYD 134 134
 FT CARBOHYD 179 179
 FT CARBOHYD 202 202
 FT CARBOHYD 252 252
 FT CARBOHYD 348 348
 FT CARBOHYD 430 430
 FT CARBOHYD 467 467
 FT CARBOHYD 576 576
 FT CARBOHYD 626 626
 FT CARBOHYD 722 722
 FT CARBOHYD 847 847
 FT CARBOHYD 856 856
 FT CARBOHYD 900 900
 FT CARBOHYD 910 910
 FT CARBOHYD 935 935
 SO SEQUENCE 1153 AA; 130471 MW; 6497BD16CE7BEC01 CRC64;
 Alignment Scores:
 Pred. No.: 1,79e-09 Length: 1153
 Score: 349.00 Matches: 259
 Percent Similarity: 34.44% Conservative: 199
 Best Local Similarity: 19.47% Mismatches: 509
 Query Match: 1.79% Indels: 363

DB: 1 Gaps: 59
 US-10-087-013-1 (1-10628) x PVDB.PLAKN (1-1153)
 QY 3010 GAAGCTAATAGAACCAAGATATGGAAGCATGAA-----TGATAT 3051
 Db 20 LysAlaAsnAsnVal---LeupheGluArgMetLysGlyIleLeuLeuLeuGluValysGlu 38
 QY 3052 ATAAATATTTGAGATTAATTCGGGACCAATCAACCAAGATGATATTCGGATAT 3111
 Db 39 AsnGluTyrValLysAsnGluInaSnGluInglYleuValaIuIcIuHis-----Tyr 56
 QY 3112 AGTATCATATACACATGTCGATGATATATCCACAAAATTAAGATGATGATGACCGAATG 3171
 Db 57 MetAspAsnAspGluIleGlu-----ArgTyrPhe---GlnGly 68
 QY 3172 GCAGATATGCTACTGCAAGGTGCGAGAAAAGAGTATGATGATGAGAGAGATGTAAT 3231
 Db 69 ThrAspArgSerArgArgValLysIleGluGluAsnValLysTyrLysTyrValaIuIcIu 88
 QY 3232 GAGTGTAAAGATTAAGATTAATGCTCAAGGCTGACGAAAGAGAGTGTACAGTTGTACG 3291
 Db 89 GluLeuAsn-----ThrLysTyrGluGluInThrLysGlyLys 100
 QY 3292 AAGTCACAGAGACCTTGTAAATGATATATATATATATATATATATATATATATATATG 3351
 Db 101 ArgIleAsnArgIleLeuLysGluSerThr-----TyrGluValaIu 114
 QY 3352 AAT 3411
 Db 115 AsnValaIaAspAsnAsnTyrIleAspAspLysAlaAsnGlyLysTyrLysThrAspAsn 134
 QY 3412 AGTGTATTAAGAGCTTCAGATACGTCACAAATATATATATATATATATATATATATAT 3471
 Db 135 -----LysThrAsnLysGlyLysGlyLysAlaAspAsnMetValMetLeu 148
 QY 3472 TTGTCGAAAT 3531
 Db 149 AspTyrAspIle-----SerGlySerGlyHisProAspGlyIleIleAspAsn 164
 QY 3532 AGTGTAT 3591
 Db 165 ValValaIuLysGlyLysThr-----GluAspGlu 173
 QY 3592 GGAATTTTGTATGATGTCACATCAAAAATGATTTTGTATGAAAAAGATGATGTAAG 3651
 Db 174 GlyAsnPheLeuGluAsnSer----- 181
 QY 3652 GATAACGAAAAATATGCTTTAGATTAACACACAGACCATGATGTCGTGTGTGT 3711
 Db 181 ----- 181
 QY 3712 AAAAGTGATGAGAACGACAGAGGTACAGATATAAAGCAAAAAAGCGGAGAAAG 3771
 Db 182 LysGlyLysAspHisProTyrArgMet-----AsnArgLys 193
 QY 3772 GATACGGAATGTAA---ACAGTGAATGATATCTTAAGAAAAAGATGGAAGAAACAA 3828
 Db 194 GluArgMetCysSerGlyValIleAsnGlnThrPheLeuIn-----LysAsnVal 210
 QY 3829 GTAGAAATATGTCATCCAAAAAGATATGATATATCCGATGTCGATGCGGAAAT 3888
 Db 211 MetArgArgGlyAsn---AsnLysArgLysArgGlyThrArgAspArgAspCysProThr 229
 QY 3889 ATAAATTTAGTGAAGACCTCGTGTGTATGATGTCGTCGAGAGACAAAAGTATATCGTA 3948
 Db 230 -----LysLysAspValCysIleProAspArgArgTyrGlnLeuCysMet 244
 QY 3949 CATTTCTGCAAT-----GATATGAAATATAAAAAA 3981
 Db 245 LysGluLeuThrAsnLeuValaAsnThrLysThrHisSerHisAsnAspIleThrPhe 264
 QY 3982 TTACAAATCAACAATTAATTAAGAAAGCTTTTCATCAATCTGCACAGACAGAAACATTTC 4041

Db 265 Leu-----LysLeuasnLeuysGluLysLeuThrTyrAspAlaIleValGluLysAsp 282
OY 4042 TTCCTACGATATATATATAAAGTAAGTACGAGAAATGACATTCGATAAAGATTA 4101
Db 283 LeuLeuLeuLysLysTyrAsnAsnValTyrSerGlu-----AspLeuLysAspIle 300
OY 4102 AAGAAGGCAAAATTCCTCCCGCATTTTGGATGCCATGTTCTACACATTTGGAGATTAT 4161
Db 301 Lys-----TyrLeuGluAspPhe 307
OY 4162 AGGATTTTATTTATGGACAGATATATCAAAAGTCATGTCAGGAGAACTAACTAAA 4221
Db 308 GlyAspIleIleMetGlyThrAspMet---GluGlyIleGlyTyrSerGluValGlu 326
OY 4222 GACCAATAGATTCCTCTTCAAAAATGTCAGCAAAAATCTCCTATGAGAAAACACGC 4281
Db 327 AsnAsnLeuArgThrValPheGlyThrGlyThrLysThrGluLeuAsp-----Arg 343
OY 4282 CAAGAATGTCGACAGACATGATGATGAGATATGGAAGCTATGCTATGTCACATGTA 4341
Db 344 LysLysTyrPheAsnGluSerLysLysTyrIleTyrGluIleThrIleLeuSerValLys 363
OY 4342 AAAATTGGGCAAAAAAAGATGATTTTACGAAACATGCTACAGT-----AACGTC 4395
Db 364 Lys-----LysLeuAsnGlyTyrSerAlaTyrPheAsnCys 374
OY 4396 AAATTTAGTCACAAAAGCACACCTTTGAGGAATTTGCCAAACGCCAGCTTTTTCGA 4455
Db 375 Lys---GluAspValGluIleAsnValGlu-----ProGluIleTyrArg 388
OY 4456 TGGCTAACCAATGTCACGACGACTATGCTATACACGACAAAATTTGGAAGATGTC 4515
Db 389 TrpIleArgGluIleTyrPheLysMetAspTyrMetSerGluLeuProLysGluIleLys 408
OY 4516 CAGGAAAAATGTAAG-----TCAAAATGACCAATGTAAGTGTATACGAA 4560
Db 409 LysGluLysCysAspArgLysLeuTyrTyrAsnLeuArgIleCysThrMetSerPro 428
OY 4561 TGTATTAAGAAATGCGAGACACTACGTTAATATATG---AAAAAAGAGATGAT 4617
Db 429 CysAsnAspSerCysLysLeuTyrAspGlnTyrIleThrArgLysLysGlnTyrPhe 448
OY 4618 CCAACAATATATATACAGATGTAACGCGACAAAAAAGATTCGATACACACAT 4677
Db 449 ValLeuSerThrLysPheSerSerValLysLysGlyGluIleLeuIleGluThrGluAsnIle 468
OY 4678 GGTATATGTTACAGACTACTGGAAGCAATGCAACAGATTAATGTAACAGGAATTT 4737
Db 469 -----ThrThrAlaTyrAspIleLeuLysGluIleLeu 479
OY 4738 ACTGCTAGTTGTGATGATAGCCCTGCAAGTGCCTGTGCTGTCACAAAGAAATATACAT 4797
Db 480 Asn-----GlyPheAsnGluValMetPheGluAsnGluIleAsn 492
OY 4798 TTGAAAAACAGCGCTTACTATGATGCCGACAAACATTTGGTCGACAAAATTTATGAA 4857
Db 493 LysArgAspAsnValTyrIleAspIle-----CysLeuLysAlaIleAspGluPro 509
OY 4858 AATGACACCAAAATATACTACATTTGAGTAAGATATGCAAAAGATTAAGTAAGAG 4917
Db 510 AsnLysAsnThrGlnGluIleLysLeuLysLysSerAlaProLysLeuGluThrGln 529
OY 4918 GCAAAACAGCGCTATTAATGAGCAAAACAAAGCTTATATACATTAATGTTGAA 4977
Db 530 ArgSerIleSerThrIleGln-----ProMetSerSerSerGlyAlaGlu 544
OY 4978 GAATGCTACAAAGATGTCCTTTCTTCCTGTCGACACTACGATATATGTTTCATGCT 5037
Db 545 LysValGlnGluLysAspLeu-----Ala 551
OY 5038 GATGCAATTAATACAGATCCAGAAATTAAGATGAAATGGTTCGAAAAAGATTGATG 5097
Db 552 HisGlyAsnIleAsnAspAlaIleTyrLysSerThrThrAsp----- 565
OY 5098 GAAGTGGCGGCAACGAGAGGATCAATTTGGTCATCTACTACAAAGAAAAA 5151
Db 566 GluAlaIleLysGlyAspGlyGluAsnGlyLysAsnGlnThrValAlaGluSerAsnIleLys 585
OY 5152 -----GAAAAAGCAAAATTAACGTCGATCGACGCAAAATAT 5190
Db 586 GlyThrAspAsnIleGluAsnGluAlaIleLysAsnValAspThrTyrLysPheValThr 605
OY 5190 ----- 5190
Db 606 GluArgSerAlaAspThrArgLysAlaThrAspIleThrGluThrGlyGluLysLeu 625
OY 5191 -----TCTATAGAGTCCCGCTTGTAGTCTAGTCAATATATGTTTATGAT 5238
Db 626 AsnThrSerTyrSerGlySerSerGluIleThrValLysLysAsnIleProGlyAspGly 645
OY 5239 -----TTAAGCATATATTTAGTATGATTAATTTGAGAGATCAAAAACAAAGCCGAG 5295
Db 646 IleValLysAspValSerAlaIleValGluAsnSerGluAsnProLeuGluThrLysHis 665
OY 5296 GAAATTTGAGAAATATTTAACAACAAATGAAACATCAGTTCGCAAGAGAT 5349
Db 666 LysIlePheGluProSerLysAspAsnSerAspAsnSerGluAsnSerGlySerMetGlu 685
OY 5350 ---GATAGTACTACAGGAAATCC-----GCTAGTACTGCGCGAAA 5388
Db 686 PheLysAlaThrSerSerAsnProIleThrGluAlaValGluSerSerAlaGluGly 705
OY 5389 TTTTTCGACCAAAATTAAGAAATGTCGTGTCGACCAATGATATGCGGCTCAAAAGT 5448
Db 706 GluValGlnGluAspSerAlaHisArgSerValAsnThr----- 718
OY 5449 GTAGGATGATGCAATTAATGCAATGCAAGATGCAAGATGATGATGATGATGATGATGAT 5499
Db 719 GlyArgAspAsn---SerThrIleSerAlaIleThrSerAspAspGlyLeuSerSerGly 737
OY 5500 ---AAAAATGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5556
Db 738 AspLysArgValGluSerIleLeuThrSerIleGluAsnAlaAspAspGlyGlyAspProVal 757
OY 5557 GAAGTACTGCTGCTATGCTTCTTCA-----TGTTCGCGCAATGGGTGAAGAT 5607
Db 758 GlnGlySerLeuThrPheAsnLeuAsnAspProSerValGlyAlaGlyGlyLysSer 777
OY 5608 TTTTGCACCAATTAAGAAAGATGAGAAATTTGTAAGGCGCTATGATGATGATGATGAT 5667
Db 778 HisIleLysThrGluGluAsnGlu----- 785
OY 5668 TGTGCTGATATGAGATTAAGAAAGAAAGATGACAGATCGTGTACAAATATAAATA 5727
Db 786 ---GlySerGlnAlaGluIleAspGlyLysAsnValAspIleAlaGluGlnIleArgThrAla 804
OY 5728 TTTATTTAGTACGAGGAACACAGAT-----GAAAAAACAATC 5766
Db 805 ThrIleThrGluValGlnProGluArgProAspLeuSerAspThrAspAsnGlyAsnVal 824
OY 5767 AAAAAATATGTCGATTAAGAAACAAATATATTTCCAGCATCCTGCGCAAAAAGATGCA 5826
Db 825 ProArgSerGly---AsnLysGln-----AsnGluLysAlaThrAlaLeuSerGly 840
OY 5827 GAGAGCGCTCGCAATATTTAGACAAACAATTAATAAATAAT 5868
Db 841 AlaGluSerLeuGluAsnGluSerValHisLysThrIleAspAsnThrThrHisGly 860
OY 5869 TGTGAAAAATTAAGTGAAGATTTGTGAATATTAAGTGTGAAGATGTCGTCACACAGCA 5928
Db 861 LeuGluAsnLysAsnGlyGlyAsnGlu-----LysAspPheGlnLysHisAsp 876
OY 5929 TTAACGTATGATATATGCAAAATATATGCGCATCATTAACGTATGCAACCAAGAGATT 5988
Db 877 PheMetAsnAsnAspMetLeuAsnAspGlnThrSer---SerAspGlnThrSerSerAsp 895


```

QY 5989 GAAGGAAGTAAATGTCAGAGTCCACGAGTCCACCGTGTACGAGGGAACACCG 6048
DB 896 GlnThrSerSerAsn--GlnThrSerSerAspGlnThrSerSerAsnGlnThrSerSer 915
QY 6049 TCACCCAGGGGATGATGATGATCAAAAGCGGCGATCGAAAAAGAGGGAACACCG 6108
DB 915 spGlnThrSerSerAspGlnThrSerSerAsp--GlnThrSerSerAspGlnThrSer 933
QY 6109 CCGGCTTACAAACACGCG--GAAAAAAGTGAAGAAATCTACAA 6147
DB 933 eSerSerAsnGlnThrSerSerAspGlnThrSerSerGlnThrSerGlnThrSerAsp 953
QY 6148 ACAGAAAT--GCGAGCA 6162
DB 953 aAlaGAsnProGluThrLeuSerSerGluAspMetSerLysGlyAspPheMetArgAsn 973
QY 6163 CAACACGAGACCGGACG--AGCAGACACAAACAGAGAAAGAGGGAACACCA 6213
DB 973 eAsnSerSerAsnGlnThrSerSerHisAsnAsnLeuAsnAsnArgLysLeuAsnThr 993
QY 6214 GCACACAAACAGAGATGACGAGGCGACAAATGTAAGCGCATCTTTCG-- 6264
DB 993 IntYrGlnHisArg-AspValLysAlaThrArgGluLysThrLeuMetSerGluVal 1012
QY 6265 AATAAACACAGATGACGAGGCGTGAATAGAGGTTGTAATCCAAACGATGACACATAT 6324
DB 1013 AsnLysCysAsnAsnArgLysSerLeuLysTyrCysAsn-- 1025
QY 6325 CCAAAATGGGGTGTATGTAGTGAAGTCAAAATGTA--AATGCGATATG 6378
DB 1026 -----ThrLeuLysAspArgMetLeuSerSerThrCys 1036
QY 6379 ATGCGCTCTAGAGGAAAAATTA--TGATTAATTAATATA 6417
DB 1037 SerArgLysThrSerLysAsnLeuCysCysSerLysSerAspPheCysLeuAsnTyrPhe 1056
QY 6418 CAATATTTAAATGTAAGTGAATGAAGCGTGAACATGATTAAGAGCGTTTAT 6477
DB 1057 GlnLeuThrProThrGluThrPheThrAsnCysMetLysGluPheGlnPheSerSerTyr 1076
QY 6478 AATATG--GCGACCAATAGAA 6495
DB 1077 GlnCysPheThrLysGlySerSerThrGlyLeuGlyLeuValTyrPheAlaThrGlyGly 1096
QY 6496 ACTCATTTTGTGTTAAATATATTAATGAATCCCGACGACGAAATGATGCA 6555
DB 1097 AlaPheLeuLeuLeuLeuLeuLeuPheValSerLysAsnValAlaSerAsnAspTyrGlu 1116
6556 AATGACAAATTCAGATGATTT 6579
DB 1117 GlnGluAlaThrPheAspGluPhe 1124
RESULT 7
PVDK_PLAYS STANDARD: PRT; 1070 AA.
AC P22290;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Duffy receptor precursor (Erythrocyte binding protein).
CN PVDK.
OS Plasmodium vivax (strain Salvador I).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=126793;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=91187056; Pubmed=1849231;
RA Fang X., Kaslow D.C., Adams J.H., Miller L.H.;
RT "Cloning of the Plasmodium vivax Duffy receptor."
RL Mol. Biochem. Parasitol. 44:125-132(1991).
CC -1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
DETERMINANT.

```

```

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: HIGH. TO P.KNOWLEDGE DUFFY RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M61095; AA63423.1;
KW Malaria; Receptor; Glycoprotein; Signal; Transmembrane.
FT SIGNAL. 1 1070
FT CHAIN. 21 1070
FT DOMAIN. 21 1007
FT TRANSMEM. 1008 1025
FT DOMAIN. 1026 1070
FT CARBOHYD. 183 183
FT CARBOHYD. 255 255
FT CARBOHYD. 351 351
FT CARBOHYD. 420 420
FT CARBOHYD. 715 715
FT CARBOHYD. 787 787
FT CARBOHYD. 825 825
FT CARBOHYD. 903 903
FT CARBOHYD. 938 938
SQ SEQUENCE 1070 AA; 119683 MW; CB051D13E294603 CRC64;

Alignment Scores:
Pred. No.: 2.9e-09 Length: 1070
Score: 344.50 Matches: 243
Percent Similarity: 32.74% Conservative: 180
Best Local Similarity: 18.81% Mismatches: 429
Query Match: 1.77% Indels: 440
DB: 1 Gaps: 60

US-10-087-013-1 (1-10628) x PVDK_PLAYS (1-1070)
QY 6918 ATGCTTCACAGATGGGAGAGAAATTTGTAATTAAGAGGAAACGTTTAAAT 6977
DB 16 LeuLeuHisLysValAsnAsnValLeuLeu-GlnTyrThrLeuThrLeuGlu-- 34
QY 6978 GAGGCGGGCTGAAGATATGAGTGAATGATGATGATGATGATGATGATGATG 7037
DB 35 -----CysLysAsnGlnTyrValLysGlyGlnAsnGlyTyrLysLeuAlaLys-- 50
QY 7038 TGCAGAGCGGTGTGAACATATCAAAATTTATTAAGAGTGGAAACTGATATGAAG 7097
DB 51 -GlyHisHisCysValGlnGluAspAsn--LeuGluTyrTrp--Le 64
QY 7098 ACAGAGAGAAAGTTCAGAAAGATTAAGATGCGCAAAAGATTAAGATTAAGATTAAG 7157
DB 64 uGlnGlyThrAsnGlnThrArgSerGlnGlnAsnLeuLysTyrLys--TyrGlyValTh 83
QY 7158 TGAAGAGACATAGAGAGGACACATGCTGCTCATGAATTTAAACATGAATTAAGA 7217
DB 83 rGlnLeuLysLysLysTyrAlaGlnMetAsnGlyLysArgSerSerArgLysLeuLysG 103
QY 7218 A-----TTATGGGCAATAG-- 7233
DB 103 uSerLysThrGlyAlaHisAsnPheGlyGlyLysSerTyrMetClnGlyLysAspGly 123
QY 7234 -GATTTGTTGTAATGCAAAACCTTCTGACACAGTACCAAAACACACACACATGACA 7292
DB 123 YAspLysThrGlyGlnGluLysAspGlyGlnHis-----LysThrAspSerLysThrAs 141
QY 7293 ATCATCCGATGATATATATGACAGATGCGTGAATTAATGTTCTGAGAAATTAACA 7352
DB 141 pAsnGlyLysGlyAlaAsnAsnLeuValMetLeuAspTyr----- 154
QY 7353 GTGTGAGTGTCTGCAACTTTCAGAAAGGATGATGATGATGATGATGATGATGATG 7412

```

Db 155 -----GluThrSerSerAsnGlyGln----- 161
 QY 7413 ACCTAAATAACCTATGATGTTGTAGAGAAAGCAGATATTATTAATCTAAAGAACAG 7472.
 Db 162 -----ProAlaGlyThrLeuAspAsnValLeuGluPheValThrGlyHisGluG 178
 QY 7473 AATAATATGATATTAACCTTGAAGAAAAATTATACCTATTGAGTCTACAAAGAAAA 7532
 Db 178 yAsnSer----- 180
 QY 7533 GGAATAATTAATAGTTGGACTAATATATCTCT-----TGGATCTCTAAGAACCTTA 7586
 Db 181 -----ArgLysAsnSerSerAsnGlyGlyAsnProTyrAspPheHisLysLysThrI 199
 QY 7587 TGCACCT-----GATTAATATATAGAGAAAGAACCCCTGTGAAATAATAGAGAGA 7637
 Db 199 eSerSerAlaIleIleAsnHisAlaPheLeuGlnAsnThrValMetLysAsnGlyAsnTyr 219
 QY 7638 AATTCGTTTAAGTATGATATATGAAATGGAATGTTACAAAAATTCAAAGTCTATACGA 7697
 Db 219 rLysAlaGlyLysArgLysGluArgLysPhePheAspCys-----AsnThr 232
 QY 7698 GAAAAAAGAGTATGTTGCTACCTCCAGAAAGAGAAACATATGCTTAAGAAATTTAGATGA 7757
 Db 232 rLysLysAspValCysIlePheAspArgLysGlnLeuGlyMetLysGluLeu----- 250
 QY 7758 AATTAATTAATGAAGCTTAAGATATATAT----- 7788
 Db 251 -ThrAsnLeuValAsnAsnThrAspPheHisArgAspIleThrPheArgLysLe 270
 QY 7789 -TATTCCTTAATAATGCTTCGTGCAACCTGACGAAATGAAGCAATGACATTAATAAAAA 7847
 Db 270 uTyrLeuLysArgLysLeuIleTyrAspAlaValAlaGluGly---AspLeuLeuLeu 289
 QY 7848 CTCAACCTCAGAGAAAGGGGCGCAATGAATCCAAATATGATAGTACTAGTAATATAGTT 7907
 Db 289 sLeuAsn-----AsnTyrArgTyrAsnLysAspPheCysLysAspIleArgTyrPheSer 307
 QY 7908 CGGTGATCTGGGTGACATAGTATAGAGAACAGATATGTCGAATGGTGGTACTATAC 7967
 Db 307 uGlyAspPheGlyAspIleIleMetGlyThrAspMetGluGlyIle---GlyTyrSerLys 326
 QY 7968 TCCCGTAAATTAATTAATATAGTTTGAATACATATATGAAAAAGCAAAATA 8027
 Db 326 sValValGluAsnAsnLeuArgSerIlePhe----- 336
 QY 8028 AATAAAGTAGAATAATTAATACAAAGATGACAAAGCTTCTGCTGGTGGGATGC 8087
 Db 337 -----GlyThrAspLeu-----LysAlaGlnGlnArgLysGlnTyrPheAsnGlu 352
 QY 8088 TAAATAGAAAGATATTTGGAAGCAATGACGTGCAAAAGCAGAGAGATGCAAAAGCTTTT 8147
 Db 352 uSerLysAlaGlnIleIlePheThrAlaMetLysSerVal----- 365
 QY 8148 TAAAAAGGAAGATGATGATTTGACAGCATTAATTAACAAGATTAAGTGTGACA 8207
 Db 366 ---LysLysArgLeuLeuLysGly----- 371
 QY 8208 TAAAGAGATCCACCTGTGATATATATA----- 8238
 Db 372 -----AsnPheIleTyrPheCysLysLeuAsnValAlaValAs 384
 QY 8239 -----CCTCAAGGTTTGGATGATGACTGAATGGTGAATATATATTGAAGCACT 8291
 Db 384 nIleGluPheGlnIleTyrArgTyrPheLeuArgLysIlePheValArgAspPheValSerGlu 404
 QY 8292 GATGGAAGATTTGAAAAATTTAAAAATGATGATGATGATGATGATGATGATGATGATG 8351
 Db 404 uProThrGluValGlnLysLeuLysGluLysCysAspGly----- 417
 QY 8352 CAAAGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8402
 Db 418 -LysIleAsnTyrThrAspLysLysValCysLysValProTyrCysGlnAsnAlaCysLys 437

QY 8403 AGAATATAAAAATTTGCTTAAATGGAATCTATTCGATATACAAATGAATAATA 8462
 Db 437 sSerTyrAspGlnTyrPheThrArgLysLysAsnGlnTyrAspValLeuSerAsnLysPhe 457
 QY 8463 CAAAGATTTGATGAACCAACCAATATACAAAAATCTGACTTATGATCATCTTCAAA 8522
 Db 457 e-----IleSerValLysAsnAlaGluLysValGlnThrAlaGlu 470
 QY 8523 TTTTGTACAAAAGTTGAAAACTTTTAAAGTGAATGCTTCTGTAAGAGCTTTCTGAATA 8582
 Db 470 yIleValIleThrProTyrAspIleLeuLysGlnGlu-----LeuAspGluPheAsnGluVal 488
 QY 8583 TCTTCATGAACAAGATTAAGTGTGATTAATTAATTAATGAAGAAATGATGCTTCCAA 8642
 Db 488 lAlaPheGluAsn-----GluIleAsnLysArgAspGly----- 499
 QY 8643 TATACGACATATGCTTTCGAAGAAACACCAAAAAGTTATTAAGAACCTTGACGTTGTAC 8702
 Db 500 -----AlaTyrIleGluLeuGlyValGlySer 508
 QY 8703 ACTACCTCTAAGATTCATTTGATTAATGCTCTACCGATCAAAACAAAGATGATGTA 8762
 Db 508 rVal-----GluGluAlaLysLysAsnThrGlu 517
 QY 8763 GGAATTAACAACCTTTTACCTTCTGCTGAGAAATGATTAATGAATATCTGATTAATG 8822
 Db 517 nGluValValThr-----AsnValAspAsnAla 526
 QY 8823 GAACGATACCTGTTCTTATATAGTTCA-----GATGATTAACAAG 8864
 Db 526 aAlaLysSerGlnAlaThrAsnSerAsnProIleSerGlnProValAspSerLysAla 546
 QY 8865 TGTATGATCTCTCCAAAGAAAGACATTTATGTAACAGCTATGACATATATTA 8924
 Db 546 aGluLysValProGlyAspSerThrHis-----GlyAsnVal 558
 QY 8925 TAAAAAGGTGATTAAGAAATTTAAAAAACTTCTTACTTCTGCTTTCAGTCAAG 8984
 Db 558 lAsnSerGlyGlnAspSerSerThrThrGlyLysAlaValThr-----GlyAspGlu 575
 QY 8985 ACAATGTTTACGCAAAATATTAATTAACGGAAGAGCTGCTTGGAGCAGATGAATA 9044
 Db 575 yGln---AsnGlyAsnGlnThrProAlaGluSerAspVal----- 587
 QY 9045 TACTTATGACAGATTTATCCGATATTAATTAAGAACATGATGATGACATCATTAATC 9104
 Db 588 -----GlnArgSerAspIleAlaGluSerValSerAlaLysAsnValAspProGlu 604
 QY 9105 TGAATAATTAATAATAATTTGAACATCAAAATGAACCAAGCAAAATGTAACATG 9164
 Db 604 nLysSerValSerLys----- 609
 QY 9165 GTGGAAAAATATACAGCTGATATGCGACGCTATGTTATGCGATATAAATTTGCTAC 9224
 Db 609 ----- 609
 QY 9225 TTCAAAAGTACATTAATGATGAGATGCTGCTCAATTACCAAGAGATGAAGAACTATAC 9284
 Db 610 -----ArgSerAspPheThrAlaSer 616
 QY 9285 GTTTCCTTGGTGTAAATGAATGGCGCAAGCAAGCATGTGAAGAAAGCAATGTAA 9344
 Db 616 r-----ValThrGlyIleAlaGluAlaGlyLysGluLysLeuLysAla 631
 QY 9345 TGATTCATTTAAAAACAATAATGCTCTGCTCAACGAAGATTAATTTGAACGCTCAGAAAT 9404
 Db 631 rAsnSer-----ArgProSerGluSerThrValGluLysAsn----- 643
 QY 9405 ATTAAGCAACCTGATGTCAGATGATTAATTAATAATATATATAGCTGATGAATATATGAT 9464
 Db 644 -----SerProGly-----AspAspThrValAsnSerAlaSerIleProValVal 658

OY	9465	AAAAATACAAAGAAAT---CTAAATATATAACCATTAAGAATCATCTTC	9521
Db	659	-----SerGlyGluAsnProLeuValThrProTyrAsnGlyLeuAlaGHisSerLysAs	676
OY	9522	AGGTAATATAGCAATAAACCATCTGAAGAA-----AATGTCACTCATATATAATC	9575
Db	676	pasnSerAspSerAspGlyProAlaGlnSerMetAlaAsnProAspSerAsnSerLysGcl	696
OY	9576	AAAAGATTCCTCAATCGGCTTTGGAGTTAAATGATATTAATGAAATAGTTACGAAACAA	9635
Db	696	YelulThrGlyLys-----GlyGlnAspAsnAspMetAlaLysAlaThrLly	711
OY	9636	AAATATGAAATAATGAAATTCAGAAAGACTACAAAAAATATATATCCGGTTATATTT	9695
Db	711	sAspSerSerAsnSerSer-----	717
OY	9696	TGTTGAGATGAAACACACAAAAAATCATGTACTAGATGGAATATTTAAAGAAAGACCA	9755
Db	718	-----AspGlyThrSerSerAlaThrGlyAs	726
OY	9756	AACAGTTCGTCTTAAGCACTATATTTCTTTACACCCCATGATGATCTTTCTATCAAGC	9815
Db	726	Pthr-----ThrAspAlaValAsp-----	732
OY	9816	ACCTTATTTCCACACATGAGATGACCAATATGATCTCTAAATATGATATTTGAAGAAG	9875
Db	732	-----	732
OY	9876	TAGTATCTCTGTTTATTTATGATATTCGCGCTTAGGTTGATAGCGCTTATTCATGAAAGA	9935
Db	732	-----	732
OY	9936	AAATATCAATACGCTGTGACACTTGTGGCTACTGAATATCCCGCAAGAGATATGG	9995
Db	733	-----ArgLuuLleAsn-----LysGcl	738
OY	9996	AATGCTACGTTGGATCCAAATAATAGTACATACATATAGAACTGGTCCATTAAGG	10055
Db	738	YAlaPro-----GluAspArgAspLysThrValGlySerLysAspGlyGlyGlyLuu--	755
OY	10056	CAAAACATATATATATATATGAGAGACATCTAGTGAGATGAAGAT-----	10105
Db	756	-----AspAsnSerAlaAsnLysAspAlaIaThrValVa	767
OY	10102	-----AAATATATGNGGACTTACCTCCCTGATATTTCTTCCTCACCAGAAAGTGAGTA	10155
Db	767	IGlyGluAspArgGlyLeuGlyLuuSerAlaGlyLysThrAsnAspArgSerLys--	786
OY	10155	TGAAGAAATGGATATATATATATATATATATGTCACAGTAGTCTCTAAATATAAACATGTAT	10214
Db	787	AsnAspThrGlyLuuLysAsnGlyAlaSerThrProAspSerLysGlnSerGluAspAlaThr	806
OY	10215	AGAAATGTACTAGAACCATCAAAAAGGAGATATACCAAGTATGATACACCAAGTAAAGA	10274
Db	806	rAlaLeuSerLysThrGluSerLeuGlnSerThrGluSerGlyAspArgThrThrAsnAs	826
OY	10275	TACACCACTAGC-----AATAGATTTATGATGATGAAAGAAAGTAACTGAAACATTA	10328
Db	826	pThrThrAsnSerLeuGluAsnLysAsnGlyGlyLysGlnLysAspLeuGlnHisLysAs	846
OY	10329	TTTGTGATGTCATATTTTCCAAATATACAGAACCAATATATATATTAACAAGTCCACATAT	10388
Db	846	pPheLysSerAsnAspThrProAsnGlnGluProAsnSerAsp--GlnThrThrAspAl	865
OY	10389	TCCAATGAATACAGAACTTAATACTTTATTTCTGATATCTCTAAGAAAAAACCTTTAT	10448
Db	865	agLu--GlyHisAspArgAspSerLleLysAsnAspLysAlaGluArgArgLysHisSye	884
OY	10449	TATATCTATTCATGATAGGGAATTTATATACTGGGAAAGAAATATGTTATATATATATAT	10508
Db	884	t-----AsnLysAspThrPheThrLysAsnThrAsnSerHisIleLuuSnnSe	900
OY	10509	GAGTACATAATTAAGTATATTCCAATGATATGCTAGAAATGATCTTATAGACGTTAT	10568

5980 AAGCAAGTTGAAGGAAGTGAATTTGTCAGTGGCCAGAGGTCCACCACTGTACGAGG 6039
563 LysGlnLysSerLysGlnSerLeu-----LysIleGlnSer 576
6040 GAAACACCGTCACCGGGGTATCTATGATCAAAACGAGCGATCGAAGAAAGAGCG 6099
577 GlnThrValAsnGlnValLysGlnValIleIleThrLeuLysSerGlnLysMetAspLeu 596
6100 AAAACAGCGCGGCTACAAAACAGCCAAAAGTGTAAATTTACACACAGAAATGCGA 6159
597 GlnSer-----ArgIleGlnLeuLysGlnLeuGln 608
6160 GCACAAACAGAAACCGACGACGACACAAACAGAAACGAAACATGACAAACAGCA 6219
609 -----GlnLeuLysThrSerValProAsn 616
6220 ACAACAGAAATCTGACGTGGGACAAATGGTAAGCCATCTTCTGAAATAACAGATAGC 6279
617 GlnAspAlaSerLysSerAsnValThrIleLysGlnLeuThrGlnLysArgAspLeu 636
6280 AGGGGTGATATGAGGGTTGTATCCAAAAGCTATGACAAATATCTTAATGGGTGT 6339
637 GlnSerGlnValGlnAsnLeuGlnThrArgIle-----SerGln 649
6340 ATTGTAGGTAGTCTAAAGAAAATGGAATGGCATATGTATGCTCTAGAGAAAAA 6399
650 IleThrArgGlnSerThrGln----- 656
6400 TTATGTATTAATATATACAAATTTAAATATGAACT-----GAAATAAG 6447
657 -----AsnMetSerLeuAsnLysGlnIleGlnAspLeuLysSerIle 672
6448 CGTGACAAATGATTAAGAGGCTTTTATTAATGTGACAAATA-----GAAACTCA 6501
673 SerAspIleSerIleLysGlnLysGlnLysSerArgIleLeuAlaGlnLys 692
6502 TTTTGTGGTTAAATATATTAATGAAAAATCTGCACAGAAATGAAATGCAAAATGCA 6561
693 PheLysLeuLeuSerAsnThrLeuAspLeuThrLysAlaGlnAsn----- 707
6562 ACAATTCACGATGATTTAAAGAAATATGATGATTAATGATGATGATTAATGATG 6621
708 -----AspGlnLeuArgLysArgPheAspTyrLeu----- 717
6622 TTTTGGAACTGATATTTCTAATGATAAAAAATTAATTAAGTAACTGAAAGTGAACA 6681
717 ----- 717
6682 ACCATTTCTCAATGAATAATTAAGAAAAACGATAAAAAAGATGAAGATTAAGT 6741
718 -----GlnAsnThrIleLeuLysGlnAspSerLysThrIleGlnThrLeuAsn 733
6742 AAAATATTTTGGAGAAAAATTAATTTTGGAGAGGATGATTAATGATTAAGTACT 6801
734 GlnThrValSerCysLysSerIleValIleGlnThrGlnLeuLeuAsnLeu--- 752
6802 TATCATCTCAGACGAAAGAAAAAGAAAAATTAAGATTAATTAAGTACTACTACT--- 6858
753 -----LysGlnGlnLysLeuArgValHisLeuGlnLysAsnLeu 766
6859 -----GACATGACCAATCTGACCGCT-----TCCCTTGAAGATTTGTAAGAG 6903
767 LysGlnGlnLeuAsnLysLeuSerProLysAspSerLeuArgIleMetValThrGln 786
6904 CCCCATTATTTAGATG-----TTCAAGAAATGGGAGAAAGATTTTGTAAAT 6951
787 LeuGlnThrLeuGlnLysGlnArgLysAspLeuGlnGlnThrArgLysSerCysGln 806
6952 AAGAGAGAGAACAGT-----TTAAATTTGAGCGCGGCTGTAAAG 6993
807 LysLysIleAspGlnLeuGlnAspAlaLeuSerGlnLeuLysLysGlnThrSerGlnLys 826
6994 GAATAT-----GAGTGTATGTAGTAAATGACGCGTAAAGACACAGAA 7035

827 AspnHisIleLysGlnLeuGlnLysAsnAsnSerAsn----- 840
7036 TGTGACAGAGCGGTGTATCAATTAATTTATTAAGAAAGTGAATATGAA 7095
841 -----IleGlnThrPyrGlnAsnLysIleGlnAlaLeuLysLysAspTyrGln 856
7096 -----AGACAAAGAGAAAGTCAAAAAGGATTAAGTGAAGTGAAGTAAAGT 7146
857 SerValIleThrSerValAspSerLysGlnThrAspIleGlnLysLeuGlnLys--- 875
7147 TATCCTTCTACGAAAGACATAGAGAAAGCAATGCTGTCATGATTTAAACATG 7206
876 ValLysSerLeuGlnLysGlnIleGlnLysAspLysIleArgLeuIleThrTyrAsnVal 895
7207 AAATTAAGAAATATGTCGAAATGAATGCTGTCTGTATGACAAAACCTTCTCAAA 7266
896 MetAspGlnThrIle-----AsnAspSerLeuArgLysGlnLeuLysSerLys 913
7267 CTACCAAAAACACACAAATCAATCAATCCGATGCTAATGATATGCGCAATCGCTG 7326
914 IleAsnLeuThrAspAlaTyrSerGlnIleLysGlnLysAspLeuThrGlnThr 933
7327 GATTAATGTTCCGAAAGATTTAACAAGTGTAGTCTGAACTTTGAAAGAGGATCT 7386
934 SerGlnSerLeuGlnGlnThrAsnSer---LysLeuAspGln---SerPheLysAspPhe 951
7387 ATGATTCATACAAAAAATTTACATGACCTTAATTAATCTATGATGAGAAAGCA 7446
952 ThrAsnGlnIleLysAsnLeuThrAspLysThrSerLeuGln-----AspLysIle 969
7447 GCATTTTATTTATCTAAAGAACAGAAAAATATATGATTAATGATTAATGATTAAT 7506
970 SerLeuLeuLysGlnGlnMetPheAsnLeuAsnGlnLysLeuAspLeuGlnLysGly 989
7507 ATACCTATGAGTCTACAAAGAAAGAAAGTAAATTAATGATGATTAATTAATCT 7566
990 MetGlnLysGlnLysAlaAspPheLysLysArgIleSerIleLeuGlnAsnAsn----- 1007
7567 TGCATCTTAAGAAACCTTATGACCTGATTAATTAATTAAGAAAGAAACCTTGTGAA 7626
1007 ----- 1007
7627 AATGAGAGAAATTCGTTTAAGTATGATTAATGAAATGTAATGAAATGAAAG 7686
1008 AsnLysGlnValGlnAlaValLysSerGlnLysThrLysSerLysIleGlnAsn 1027
7687 TTCTATCAGAGAAAAAAGAGTATGTACCTCCAGAGAGAGAAATGATG-----TG 7739
1027 AspLeuAspGlnGlnThrIleThrAlaSerThrAlaGlnAsnAsnThrGlnGlnIle 1047
7740 CTTAAGCAATTAAGATTAATAA-----ATTGAAAGACTTAAGATTAATTAATCTCT 7796
1047 GlnLysHisIleAspValSerLysThrIleSerGlnLeuArgGlnLeuHisThrTyr 1067
7797 AAAATGTTCTGTCGACCTGACCAATGAAAGCAATGACATTAATAAATCAATCTCA 7856
1067 TyrGlnGlnValLysThrLeu---AsnLeuSerArgAspGlnLeuLysAlaLeuLys 1086
7857 AGAGAAC-----GGTGGCAATGAATCCAAATGTGATGATCTGAATATAG 7904
1086 GlnLysGlnLysSerThrPheSerGlnLysGlnSerLeuGlnGlnLeuAspLeu 1106
7905 TTTCGCTGATCTGGGTGACATTAATGAGAACAGATATGATTAAGATTTGTTACTT 7964
1106 ThrSerArgIleGlnAspLeuSerSerGlnAsnLysLeuLys----- 1121
7965 ACCCTCCGTAGAAATTAATATATAAGTTTGAATATATATGAAATGAGAAA 8024
1122 -----AspGlnIleGlnIleThrAlaAlaAsp-----LysGlnValAsnAsn 1136
8025 TAAAAATTAAGTGAATTAATATACATGATGACAAAGCTTCTGTTGCTTGGTGGGA 8084

QY 10088 GTGAGATGAGATGATATATATGCTGACTATCTCTCTGATATATCTATCCGAAA 10147
 Db 1763 snileProLaserArglyLeu-----lleSerSerSerThrleuSerThrAspT 1781
 QY 10148 GTGAGATGAGATGATGATATATGATATATATGATACAGGTGCTCTATATATAA 10207
 Db 1781 hrAsnAspLueLueLueLueLueLueLueLueLueLueLueLueLueLueLueLue 1801
 QY 10208 CATTGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10251
 Db 1801 aLlSerSerLueLueLueLueLueLueLueLueLueLueLueLueLueLueLueLue 1821
 QY 10252 -----AGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10306
 Db 1821 laileLueLueLueLueLueLueLueLueLueLueLueLueLueLueLueLue 1835
 QY 10307 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10366
 Db 1835 luvalGlyLueLueLueLueLueLueLueLueLueLueLueLueLueLueLueLue 1850
 10367 ATAAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10426
 1850 leasn-----GluSerLysLysIleLysThrGlu 1860
 QY 10427 ATCTGAGAGAAA 10440
 Db 1860 spgluglulLys 1864
 PVDG_PLAKN
 ID PVDG_PLAKN STANDARD; PRT; 1070 AA.
 AC P50494;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Duffey receptor, gamma form precursor (Erythrocyte binding protein).
 OS Plasmodium knowlesi.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 ON NCBI_TaxID=5850;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92357776; PubMed=1496004;
 RA Adams J.H., Sim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.;
 RT "A family of erythrocyte binding proteins of malaria parasites.",
 RL Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).
 CC -1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFEY BLOOD GROUP
 DETERMINANT.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: HIGH. TO P.VIVAX DUFFEY RECEPTOR.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

 DR EMBL; M90695; AAA29604.1;
 KW Malaria; Receptor; Glycoprotein; Signal; Transmembrane;
 KM Multigene family.
 FT SIGNAL 1 21
 FT CHAIN 22 1070
 FT DOMAIN 22 1003
 FT TRANSMEM 1004 1025
 FT DOMAIN 1026 1070
 FT CARBOHYD 134 134
 FT CARBOHYD 139 179
 FT CARBOHYD 676 676
 FT CARBOHYD 743 743
 FT CARBOHYD 785 785
 FT CARBOHYD 936 936
 FT SITE 279 281
 POTENTIAL.
 DUFFEY RECEPTOR, GAMMA FORM.
 EXTRACELLULAR (POTENTIAL).
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CELL ATTACHMENT SITE (POTENTIAL).

SEQ SEQUENCE 1070 AA; 120931 MW; 703D68811BC11B50 CRC64;
 Alignment Scores:
 Pred. No.: 6.14e-09
 Score: 337.50
 Percent Similarity: 33.33%
 Best Local Similarity: 18.59%
 Query Match: 1.73%
 DB: 1
 Gaps: 60
 US-10-087-013-1 (1-10628) x PVDG_PLAKN (1-1070)
 QY 3040 ATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3099
 Db 35 LeuGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 54
 QY 3100 TATGCGGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3159
 Db 55 His-----TyrMetAspAsnAspGlnIleGlu-----ArgTyr 65
 QY 3160 ATGACCGGATGCGCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3219
 Db 66 Leu-----GlnGlyThr 69
 QY 3220 GAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3279
 Db 70 AspArgSerArgArgValLysIleGluLysLysLysLysLysLysLysLysLysLys 80
 QY 3280 ACAGGTGTCAGACAGTGCACAGACGCTGTAATGATGATGATGATGATGATGATGATG 3339
 Db 81 -----Val 81
 QY 3340 AAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3393
 Db 82 LysTyrLysTyrAsnValGluLysLysLysLysLysLysLysLysLysLysLysLysLys 100
 QY 3394 CAATGCTGTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3453
 Db 101 -----ArgIleAsnArgIleLysLysLysLysLysLysLysLysLysLysLysLys 118
 QY 3454 AGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3513
 Db 119 AsnAsnTyrIleAsnAspLysLysLysLysLysLysLysLysLysLysLysLysLys 138
 QY 3514 -----AGTGTACTAGTGAAGA 3531
 Db 139 GlyGluGlyAlaArgAsnMetValMetLeuAspTyrAspIleSerGlyGlnPro 158
 QY 3532 AGTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3588
 Db 159 AspIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 172
 QY 3589 ACAGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3648
 Db 173 GluLysAsnHe-----LeuLysAsnSerSerLys 182
 QY 3649 AAGATACGAAATATGCTTTAGATTAACACAGAGCAGCATGATGCTGCTGCTGCTGCT 3708
 Db 183 GlyAspAspHisProTyrArgMetLysArgLysGluLysMetSerSerGlyAla 200
 QY 3709 TGTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3768
 Db 200 ----- 200
 QY 3769 AAGATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3828
 Db 201 -----IleAsnGlnIlePheLeuGlnAsnAsnValMetAspLys 213
 QY 3829 GTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3888
 Db 214 CysAsnAsp-----LysArgLysArgGlyGluArgAspTyrAspPysProThr 229
 QY 3889 ATAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3948

Db 230 -----GluLysAspValCysIleProAspArgArgGlnIleCysMet 244
OY 3949 CATTTCCTGGCAAAAT-----GATTAATGAAATATAA 3978
Db 245 MetGlnIleThrAsnLeuValAspThrAspThrHisPheHisSerAspIleIlePheArg 264
OY 3979 AAATTACATCACAAAGTTAATTA-----AAAGAGCTTTCATC 4017
Db 265 LysSerThrSerArgArgLeuIleTyrAspValGlyGlyArgGlyAspIleLeu 284
OY 4018 AAATCTGCACAGAGAAACATTCTTCTCATGATATATATAAATAAGATGAGGAA 4077
Db 285 Lys-----LysTyrAsnAsnValTyr 291
OY 4078 GGAATAGACATCGATTAAGCAATTAAGAGCAAAATCTCCCGCATTTTTCAGATCC 4137
Db 292 SerGlnAspLeucylsAspIleLys----- 300
OY 4138 ATGTTCTACACATTGGAGATTATAGCATTTTATTTGAGACAGATATATCAAAAGT 4197
Db 301 -----TrpSerLeuGlnAspPheGlyAspIleIleMetGlyThrAspMet---GluGly 317
OY 4198 CATGTGAGGAGTAACATAAAGACCAATAGATTCTCTTTCAAAATGCGTGCACAA 4257
Db 318 IleGlyTyrSerLeuValValGlnAsnAsnLeuArgSerIlePheGlyThrSer 337
OY 4258 AAATCTCTAATGAAAAAACAAGCAAGATGTGTGACAGACATAGTATGATATG 4317
Db 338 AlaGlnLeuAsp-----ArgLysLysThrTrpAsnAspHisLysLysAspIleTrp 354
OY 4318 GAAGCTATGCTATGTGCACATGTAATAATGGGCAAAAAAGATGATTTTACCGAAAC 4377
Db 355 LysIleMetIleLeuSerValLysGlu-----LysAsnArg 366
OY 4378 TAGCGTTACCAACACGTAATTTAGTACAAAGACCACTTGGAGATTTGCCAA 4437
Db 367 TyrSerAlaTrpAsnCysLys---GluAspValGlnIleLysValGlu----- 381
OY 4438 CGAAGCCAGTTTTTACGATGGCTAACCGAATGTGTACGACGACATATTCATATACGACA 4497
Db 382 ---ProGlnIleTyrArgTrpIleArgGlnTrpLysArgAspTyrMetSerGlnPheArg 400
OY 4498 AAATATTGAAGATGTGCAGAAAAATGTAAGTCAATGACCAATGAGTGCATACA 4557
Db 401 GlnGlnIleArgLysLeuAsnGlnLysCysGlu---AspLysLeuTyrTyrSerThr 418
OY 4558 -----GAATGTAATTAAGAAATGCGGAGCTACGTTAAATATATG 4596
Db 419 MetLeuIleCysThrLeuProProCysAsnAsnAlaCysLysSerTyrAspGlnTrpIle 438
OY 4597 AAA---AAAAAAAAGAGTGATTCACAGATTAATTAATTAAGATGAACGCGACAAA 4653
Db 439 ThrGlyLysLysLysGlnTrpAspValLeuSerThrLysPheSerSerValLysLysAla 458
OY 4654 AAAAGATTCCATACACACACATTGCT-----GTAATGTTTACGACTATATCT 4701
Db 459 GlnLysIleGlnTrpGlnAsnIleAlaArgLysAspIleLeuLysGlnIleLeuAsn 478
OY 4702 GGAACGATGACACAGATTAATGACAGAAATTTACTGATGTTGGTGATTAAGCT 4761
Db 479 GlyPheAsnGlnValIleThrPheGlnAsnGlu----- 488
OY 4762 GGAAGTGCTCTGTGTGTAACAAGAAATATTAATTTAGAAAAACAGGCTTACTATGAT 4821
Db 489 -----IleAsnLysArgAspLysLeuTyrAsn-----TyrPhe--- 499
OY 4822 CCGCAACAACATTGTGGTGCACAAAATTTATGAAAATGACGCAAAATATACATACATT 4881
Db 500 -----CysValCysIleValGlnGlnLysLysAsnThrGlnGln 513
OY 4882 TCGAGTAAGATTAAGTCAAAAGATTTGTAAGAGGCAAAACAGAGGCTATTAAGTGG 4941
::: |||||

Db 514 Asn-----ValLysAsnValGlySerGlyVal----- 522
OY 4942 CAAAACAAGGCTCTTAATTAACATTAATTAATTAAGAAATGACTGAAGATGCTTTT 5001
Db 523 GluSerLysAlaProSerSer---AsnProIleAsnGlnAlaValLysSerSerGly 541
OY 5002 CCTTCGTGCTGACATCACTATATGTTTCAT-----GCATTG 5077
Db 542 GlnGlyLysValGlnGlnAspSerAlaHisArgSerValAsnGlnGlyGlnGlyLysSer 561
OY 5038 GATGCAATTAATTAACATTAACATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 5097
Db 562 SerThrAsnGlnAlaAspProGlySerGlnProGlyLysProLysAspArgSerValAsp 581
OY 5098 GAAGTGGCGCGACAGGAGGATGCAATTTGGTCAATTAATTAACAAAGAAAAAGAAAA 5157
Db 582 GlnLysAlaGlyValProLysLeuSerAlaGlyGlnGlnLysAspLysValProProLys 601
OY 5158 GAGAAATTAATAAGCTGGATGCGCACAAATTTCTATGAGTCCCGCTTGTATGCT 5217
Db 602 GlnAlaAlaIleThrGlnSerAlaValProHisSerAlaAspLysThrProIleThrAla 621
OY 5218 ATGAATATAGTTTATGATTTAAGATTAATTAATTAATTAATTAATTAATTAATTAAGAT 5277
Db 622 -----ThrGlnGlu 624
OY 5278 GAAAAACAAGGACGAGAAATTTTGAAGAAATTAATTAACAAATTAATTAATTAATTAAT 5337
Db 625 AsnLysGlnArgThrGln-----ValAspGlyValAlaGlyLysArgGlyLysAlaPro 642
OY 5338 GGCAAA-----GGAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5367
Db 643 GlyProThrValSerSerAspValProSerValGlyGlyLysAspSerGlyProSerThr 662
OY 5368 CCGGATAGTACGCGCGCAAAATTTTTCGAACGAAATTAAGATGATGTTGTGACGCA 5427
Db 663 ProLysSer-----HisLeuAlaGlyLysAsnGlyGlu---ValHisAsnGly 677
OY 5428 ATGATATGCGGTACCAACGTTGATGAGTATGAAATTAATTAATTAATTAATTAATTAAT 5487
Db 678 -----ThrAspThrGluProLysGlnAspGlyGlnLysAlaAspPro 691
OY 5488 GATGAGATCTAAAAAATGCGTTCTGTACCTCAATGAT-----GATTATTCATG 5541
Db 692 GlnLysAsnIleGlnValLysGlyLysGlnAspThrAspAspArgSerGlnLysSerLeu 711
OY 5542 GGGAAAAATCCGATGAAGTACGCTGATCAAGTTCTTCGATGCTTTCGATGCTTTCGATGCT 5601
Db 712 GlyProHisThrAspGlnArg-----AlaSerLeuGly 722
OY 5602 GAAGATTTTTCCAACATTAAGAAAGAAATGCGAAATTTGTAAGGCGCTG--- 5655
Db 723 Gln-----ThrHisMetGlnLysAspThrGlnTrpThrGlyLysSerThrLeuThr 739
OY 5656 -----AATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5697
Db 740 ProGlnGlnAsnValSerValAlaSerAspAsnGlnValProGlySerGlyAsnLys 759
OY 5698 TGTAACAGATGCGGTGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5757
Db 760 GlnAsnGlnGlyAlaThrAlaLeu----- 767
OY 5758 AAACAAATCAAAAAATATGATGATTAAGAAAGAAATTAATTAATTAATTAATTAATTAAT 5817
Db 768 -----SerGlyLysGlnSerLeuGlnSerSerGlySerValHis 780
OY 5818 AAAGTGCAGAGAGCGTCCGGAATTTTATGACAAACAATTAATTAATTAATTAATTAAT 5877
Db 781 LysThrIleAspAsnThrThrHisGlyLeu-----GlnAsn 792
OY 5878 AAAAGTGAGATTTGATATATAGTATTAAGATGTAAGATGTCACACAGCATTAATGAT 5937
Db 793 LysAsnGly----- 795


```

Db 624 -----ThrgluGlu-----AsnlysgluGlyThrGlnMetAsp 634
QY 5332 -----TCA 5334
Db 635 G1yAlaIaIaGlyLysAspGlyLysAlaIaProGlyProThrValSerSerAspValProSer 654
QY 5335 GTTGGCAAGAAAGTAGTACTACTACAGAAATCCCGGTAGTACTGCGCGGAAATTTTTC 5394
Db 655 ValGly---GlyLysAspSer-----GlyProSerThrSerAlaSerHisAlaLeu 670
QY 5395 TGGACGAAATAAGAAATGTGTGTGAGCAATGATATGCGGTGACAAACGTGTAG 5454
Db 671 AlAGlyGluAsnGlyGlu---ValHisAsnGly-----ThrAspThr 683
QY 5455 GATGATGCAAGAAATAGTCAAAATAGTCAAGAAATGATCAATCAAAATGATGTTCT 5514
Db 684 GluProLysGluAspGlyGlyLysAlaAspProGlnLysAspIleGluValLysGlyLys 703
QY 5515 GTACCTTCAGATGAT-----GATTATCCTATGCGGAAATCGCATGAAAGGTACTGCG 5568
Db 704 GluAspThrAspAspArgSerGlnGlySerIleGlyProHisThrAspGluArg----- 721
QY 5569 TATCAGTTCTCTGATGATGTTGCGCAATGGCGTGAAGATTTTTCGAAACATTAAGAAAG 5628
Db 722 -----AlaThrLeuGlyGlu-----ThrHisMetGlyLys 731
QY 5629 GAATTCGAGAAATGTTGGTAGGGCGGT-----AATGATTATACTTGTGCT 5673
Db 732 AspThrGluThrAlaGlySerThrLeuThrProGlnGlnAsnValSerValAlaSer 751
QY 5674 GATTAATGAGAT-----AAAGAAAGAAATGTCAGATCGCTGTACACATTAATAA 5724
Db 752 AspAsnGlyLysAsnValProGlySerGlyAsnLysGlnAsnGlnGlyAlaThrAlaLeu--- 770
QY 5725 AATTTATTAGTAGAGTGAAGAACACAGTATGAAAAAATCAATCAAAATATGTTGAGAT 5784
Db 771 -----SerGly 772
QY 5785 AAAGCAAAATATATTCACAGATCTCTGCGAAAGATGCGAGACCGTCCGCAATAT 5844
Db 773 AlaGluSerLeuLysSerAsnGluSerValHisLysThrIleAspAsnThrThrHisGly 792
QY 5845 TTAGACAAACAATTAATAAATAATTTGTGAATAAAGAGATTTGTAATTAATTAAGT 5904
Db 793 Leu-----GluAsnLysAsnGlyLysAsnLys----- 801
QY 5905 ATGAAAGATGTGTCCACACAGGATTAAGTGAATGTAATGTAATGTAATGTAATGTAATG 5964
Db 802 ---LysAspPheGlnLysHisAspPheMetLysAsnAspMetLeuAsnAspIleAlaSer 820
QY 5965 TTAGACGATGACCAAAAGAAAGTGAAGAAAGTGAATTTGTCAGATGCCAGAGTCCA 6024
Db 821 SerAspHis---ThrSer-----SerAspGlnThrSerSera 832
QY 6025 CCACGTGTACGAAAGGAAACACCGTACACAGGATATCATCATATCAAAAGCGACGCA 6084
Db 832 spHisThrSerSerAspGlnThrSerSerAspHisThrSerSerAspHisThrSerSera 852
QY 6085 TCGAAAAAAGAAAGCAAAACAGCGCGCTACAAAACAGCGGAAAAAGTGA-----A 6138
Db 852 spGlnThrSerSerAspGlnThrSerSerAspGlnThrIleAspThrGlnGlyHisHis 872
QY 6139 AATTAACAAACGAAATGCGAGCAACAAACAGCAACCGGCAAGCGCAACAAACAGCA 6198
Db 872 rGAspAsnValAlaTrpAsnProGluLysLysSerGluAspMetSerLysGlyAspPhe 892
QY 6199 AAACGACATCAACAGCAAC-----AACCAAGATCT--- 6231
Db 892 eTArgAsnSerAsnSerAsnGlnLysLysSerHisAsnAsnLeuAsnAsnArg-LysLeu 911
QY 6232 -----GACGTG---GGCACAAATGTAAGGCAATCTTTTCG 6264

```

```

Db 912 AsnArgAspGlnThrGlnHisArgAspValLysAlaThrArgGluLysIleIleLeuMet 931
QY 6265 -----AATTAACCAATACAGAGGTGAAATAGAGGTTGTAATCCAAAAACGTAT 6315
Db 932 SerGluValAlaAsnLysCysAsnAsnArgAlaSerValLysLysCysAsn----- 947
QY 6316 GGACATATCTCAATTAATGGGTTGATTGTAGTAAGTCTCAAGAAATGAA-----AAT 6369
Db 948 -----ThrIleGluAspArgMetLeuSer 955
QY 6370 GGCATATGATGCTCTCTAGAGAAAAAATTA-----TGTAATA 6408
Db 956 SerThrCysSerArgGluAlaTrpAlaGlyAsnLeuLysCysSerIleSerAspPheCysLeu 975
QY 6409 AATTAATATCAATATTAATTAATGAAATGAAATAGCGTGAACATGATTAATAAGAG 6468
Db 976 AsnTrpPheGlnLeuLysSerGlyGluPheTrpAsnCysMetLysGluGlnPheGluAsp 995
QY 6469 GCTTTATTAATATG-----GACGCAATAGAA 6495
Db 996 ProSerTyrGluCysPheThrLysGlySerThrGlyIleValTyrPheAlaThrGly 1015
QY 6496 ACTCAATTTTGTGTTAAATATATATTAATGAAAT-----CCTGCAGACAAATGAAATG 6552
Db 1016 G1yAlaPheLeuIleIleLeuLeuLeuPheAlaSerTrpAsnAlaAlaSerAsnAspTyr 1035
QY 6553 CAAATATGGAACAATTCAGATGATTT 6579
Db 1036 GluGluGluAlaThrPheAspGluPhe 1044

RESULT 11
MS2_DICDI
ID MS2_DICDI STANDARD; PRT: 2116 AA.
AC P08799;
DI 01-NOV-1988 (Rel. 09, Created)
DI 01-OCT-1989 (Rel. 12, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin II heavy chain, non muscle.
GN MHCA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RP MEDLINE=87092266; PubMed=3540939;
RA Warrick H.M., de Lozanne A., Leitwand L.A., Spudis J.A.;
RT "Conserved protein domains in a myosin heavy chain gene from Dictyostelium discoideum."
RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RP STRAIN-AX2;
RC MEDLINE=90353583; PubMed=2387408;
RA Luck-Vielmeier D., Schleicher K., Grabatin B., Wippler J.,
RA Getts G.;
RT "Replacement of threonine residues by serine and alanine in a phosphorylatable heavy chain fragment of Dictyostelium myosin II."
RL FEBS Lett. 269:239-243(1990).
RN [3]
RP PHOSPHORYLATION SITES.
RP MEDLINE=88112226; PubMed=2828113;
RA Wagle G., Noegel A., Scheel U., Getts G.;
RT "Phosphorylation of threonine residues on cloned fragments of the Dictyostelium myosin heavy chain."
RL FEBS Lett. 227:71-75(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RP MEDLINE=95345066; PubMed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Holden H.M.,
RA Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium discoideum complexed with MgADP, BeFx and MgADP.AlF4-."
RL Biochemistry 34:8960-8972(1995).

```

RN [5] X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
 RP MEDLINE-95345067; PubMed-7619796;
 RA Smith C.A., Rayment I.;
 RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
 RT truncated head of Dictyostelium discoidium myosin to 2.7-A
 RT resolution.";
 RL Biochemistry 34:8973-8981(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
 RA MEDLINE-96206189; PubMed-8611530;
 RT Smith C.A., Rayment I.;
 RT "X-ray structure of the magnesium(II)-ADP vanadate complex of the
 RT Dictyostelium discoidium myosin motor domain to 1.9-A resolution.";
 RL Biochemistry 35:5404-5417(1996).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
 RA MEDLINE-97452580; PubMed-9305951;
 RT Gullick A.M., Bauer C.B., Rhoden J.B., Rayment I.;
 RT "X-ray structures of the MgADP, MgATPgammaS, and MgAMPNP complexes
 RT of the Dictyostelium discoidium myosin motor domain.";
 RL Biochemistry 36:11619-11628(1997).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
 RA MEDLINE-98070605; PubMed-9405148;
 RT Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
 RT "X-ray crystal structure and solution fluorescence characterization
 RT of Mg.2(3+)-O-(N-methylanthraniloyl) nucleotides bound to the
 RT Dictyostelium discoidium myosin motor domain.";
 RL J. Mol. Biol. 274:394-407(1997).
 CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
 CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
 CC -1- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
 CC INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
 CC LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
 CC (MLC-2).
 CC -1- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
 CC CORTEX.
 CC -1- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMN) AND 1 HEAVY MEROMYOSIN (HMN). IT CAN BE FURTHER
 CC SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (SI) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -1- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
 CC THE ACTIN-ACTIVATED ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE
 CC ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
 CC POSITION (688).
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: M14628; AAA33227.1; -
 DR PIR; A26655; A26655.
 DR PIR; S00250; S00250.
 DR PDB; 1MNA; 03-DEC-97.
 DR PDB; 1MND; 17-AUG-96.
 DR PDB; 1MNG; 03-DEC-97.
 DR PDB; 1MNI; 03-DEC-97.
 DR PDB; 1MND; 17-AUG-96.
 DR PDB; 1MNE; 17-AUG-96.
 DR PDB; 1VOM; 23-DEC-96.
 DR PDB; 1LVK; 28-JAN-98.
 DR Dictydb; DD01008; mhca.

DR InterPro: IPR000048; IQ region.
 DR InterPro: IPR004009; Myosin_N.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00063; myosin_head; 1.
 DR Pfam: PF00612; IQ; 2.
 DR Pfam: PF02736; Myosin_N; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR PRODOM: PD000355; myosin_head; 1.
 DR SMART: SM00015; IQ; 1.
 DR SMART: SM00242; MYSC; 1.
 DR PROSITE: PS50096; IQ; 1.
 KW Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
 KW Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
 FT DOMAIN 1 761
 FT 762 791
 FT 817 2116
 FT 179 186
 FT 638 660
 FT 738 752
 FT 130 130
 FT 678 678
 FT 1823 1823
 FT 1833 1833
 FT 2029 2029
 FT MOD_RES 2029 2029
 FT MOD_RES 2029 2029
 SQ SEQUENCE 2116 AA; 243871 MW; 2FC3770BB1EE56A1 CRC64;
 Alignment Scores:
 Pred. No.: 1,53e-08 Length: 2116
 Score: 329.00 Matches: 459
 Percent Similarity: 33.08% Conservative: 381
 Best Local Similarity: 18.08% Mismatches: 958
 Query Match: 1.69% Indels: 741
 DB: 1 Gaps: 103
 US-10-087-013-1 (1-10628) x MYSL2.DICDI (1-2116)
 QY 556 AATCATCCATAAAGAGACGCTTCGACGCTTCGACGAGTTTCAGAT 615
 DB 2 AATCATCCATAAAGAGACGCTTCGACGCTTCGACGAGTTTCAGAT 615
 QY 616 ATAGGTATATTTAGAGAGATATGATATGATATGATATGATATGATAT 669
 DB 2 ATAGGTATATTTAGAGAGATATGATATGATATGATATGATATGATAT 669
 QY 22 -----SerAspLeuPheIsthrValSerAspLeuPheIsthrVal 34
 DB 22 -----SerAspLeuPheIsthrValSerAspLeuPheIsthrVal 34
 QY 670 -----GTA 672
 DB 670 -----GTA 672
 QY 35 IletPtyrAsnProAspProLysGluArgSerPheTyrGluCysGluIleValSer 54
 DB 35 IletPtyrAsnProAspProLysGluArgSerPheTyrGluCysGluIleValSer 54
 QY 673 GAACGGGCTCCGACAGCTTTCAAGAAATACATATGATATGATATGATATGAT 732
 DB 673 GAACGGGCTCCGACAGCTTTCAAGAAATACATATGATATGATATGATATGAT 732
 QY 55 GlutThrSerAspSerPheThrPheIsthrVal-----AspGlyGlnAspArgGlnValLys 73
 DB 55 GlutThrSerAspSerPheThrPheIsthrVal-----AspGlyGlnAspArgGlnValLys 73
 QY 733 -----AATGATTCATCAATCT-----GATGATCTGCAATATATATATAA 771
 DB 733 -----AATGATTCATCAATCT-----GATGATCTGCAATATATATATAA 771
 QY 74 LysAspAspAlaAsnGlnArgAsnProIleLysPheAspGlyValGlnAspMet----- 91
 DB 74 LysAspAspAlaAsnGlnArgAsnProIleLysPheAspGlyValGlnAspMet----- 91
 QY 772 TTAAGACAGCATGCTGCAATGCAATGAATGAATGAATGAATGAATGAATGAT 831
 DB 772 TTAAGACAGCATGCTGCAATGCAATGAATGAATGAATGAATGAATGAATGAT 831
 QY 92 -----SerGluLeuSerTyrLeuAsnGlnIuproAlaValLys 103
 DB 92 -----SerGluLeuSerTyrLeuAsnGlnIuproAlaValLys 103
 QY 832 GCATCATATAATCTGATATTTATGCAATGCAAGATATATACCATTTATTTCAAT 891
 DB 832 GCATCATATAATCTGATATTTATGCAATGCAAGATATATACCATTTATTTCAAT 891
 QY 103 eHis-----AsnLeuArgValArgTyrAsnGlnAspLeuIleTyrThrSerGlyLe 121
 DB 103 eHis-----AsnLeuArgValArgTyrAsnGlnAspLeuIleTyrThrSerGlyLe 121
 QY 892 CCTAATGCGGCATTAACAAGAGAGTTCTCAATTTAGATTATGATGTCCTC---AA 948
 DB 892 CCTAATGCGGCATTAACAAGAGAGTTCTCAATTTAGATTATGATGTCCTC---AA 948
 QY 121 uPheLeuValAlaValAsnProPheLysArgIleProIleTyrThrGlnGluMetValAs 141
 DB 121 uPheLeuValAlaValAsnProPheLysArgIleProIleTyrThrGlnGluMetValAs 141
 QY 949 TATTTACGTTGCTTCGACGAATGGGAGACAGACT-----TTTCCGCAAAAAGAAAT 999
 DB 949 TATTTACGTTGCTTCGACGAATGGGAGACAGACT-----TTTCCGCAAAAAGAAAT 999
 QY 141 pIlePheLysGlyArgArgArgAsnGluValAlaProHisIlePheAlaIleSerAsp-- 160
 DB 141 pIlePheLysGlyArgArgArgAsnGluValAlaProHisIlePheAlaIleSerAsp-- 160
 QY 1000 ATTAATGTAAGAAAGGTCAAGAGACTCTGCTGATATGCAAGAAAGCGTTATTTGATGT 1059
 DB 1000 ATTAATGTAAGAAAGGTCAAGAGACTCTGCTGATATGCAAGAAAGCGTTATTTGATGT 1059

Db 161 ValAlaIYrArgSerMetLeuAspasp---ArgGlnAsnGlnSerLeuLeuIleThrgly 179
 1060 CATAATGACATGATTTGACGACACATTTGCAAAAAAGATTTTCAT----- 1110
 Db 180 GluSerGlyAlaGlyLysThrgLysnThr---LysLysValIleGlnIYrLeuAlaSer 198
 1111 -----TTGATTAATAGTACTAGCTGT 1134
 Db 199 ValAlaGlyArgAsnGlnAlaAsnGlySerGlyValIleGlnGlnIleLeuGlnAla 218
 1135 TCGACTAATGCAAGTTTGGAACTTTGTTAGGCAATCAACAAGACATTTAAAAA 1194
 Db 219 AsnProIleLeuAlaIlePhe-----GlyAsnAlaLysThrThrArgAsnAsn 234
 1195 CAAAAAGAAAAATATGAAAA----- 1215
 Db 225 AsnSerSerArgPheGlyLysPheIleGlnIleGlnPheAsnAlaGlyPheIleSer 254
 1216 -----GAAATACATCATTTATTCGAAACGATTAACAATTTGCAATTAATTAAGT 1269
 255 GlyAlaSerIleGlnSerTyrLeuLeuGlnLysSerArgValValPheGlnSerGluThr 274
 1270 GAA---TATTATAACATTTTATGAAAACTT-----AAGGA 1305
 Db 275 GluArgAsnTyrHisIlePheTyrGlnLeuLeuAlaGlyAlaThrAlaGlnLysLys 294
 1306 ACGCAATATGCACTAATGACACTTTTAAATTTACTAATTAAGAAAGATTTGTA 1365
 Db 295 AlaLeuHisLeuAlaGlyProGlnSerPheAsnTyrLeuAsnGlnSerGlyTyrValAsp 314
 1366 ---GGAGATTAACGAGGAAAAAGATTTACTTTACTAACGCTGCTGACAAAGG 1422
 Db 315 IleLysGlnIleLysArgSerSerGlnLysPheLysIleThrArgGlnAlaMetSerPheVal 334
 1423 ATATTTCGTGTCGATATTTGCGCAAGTGTCCGACCTGGGG-----GTCAATGT 1476
 Db 335 GlyPheSerGlnGlnGlnMetSerIlePheLysIleIleAlaGlyIleLeuHisLeu 354
 1477 GATGTTATAAATACACACACAAATCAGAT-----AATGATCGTAACGT 1521
 Db 355 GlyAsnIleLysPheGlnLysGlyAlaGlyGlnGlyAlaValLeuLysAspThrAla 374
 1522 GTAATATGAAAGCTAATTAACCTCAGGGGTGGAAGCCCTAATATCTCTC--- 1578
 Db 375 LeuAsnAlaAla---SerThrValPheGlyValAsnProSerValLeuLysAla 392
 1579 -----CTTTATAGTGTATGTAAGCAAGGTGATATTAACAACAAATTAAGA 1623
 393 LeuMetGlnProArgIleLeuAlaGlyArgAspLeu---ValAlaGlnHisLeuAsn 410
 1624 AATTTTGTAACTGCTAATAT----- 1647
 Db 411 ValGlnLysSerSerSerArgAspAlaLeuValLysAlaLeuTyrGlyArgLeuPhe 430
 1648 -----TCAAAAGATAAAAATAT-----CAAAATGGGAATGCTATAT--- 1686
 Db 431 LeuTyrLeuValLysLysIleAsnAsnValLeuGlnGlnLysAlaGlyValPheIle 450
 1687 -----AAGATGAAATATTAATAGATGTAACCTGGAACAAATCTGTA 1731
 Db 451 GlyValLeuAspIleSerGlyPheGlnIlePheLysValAsnSerPheGlnLeuLys 470
 1732 ATCAATATGATTAATCTAATATATATAT-----CATATTTTTCATTAATG 1785
 Db 471 IleAsnTyrThrAsnGlnLysLeuGlnGlnPhePheAsnHisHisMetPheLysLeuGln 490
 1786 GTTACATATTTATGAGGATATTAAGTGAATGACAAACTTAACCTGTATA--- 1842
 Db 491 GlnGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 505
 1843 -----AATTAATACACACGATGTGATTTGATGATTAACAGAAATTCCTTA 1890

Db 506 PheGlyLeuAspSerGlnAlaThrIleLeuAspLeuIleAsp----- 518
 1891 TGTTTGACATGGGTTAAACAAAAGAGAAATGGAATGATTAAGAACTGTTC 1950
 Db 519 -----GlyArgGlnProProGlyIleLeuAlaLeu 529
 1951 ACAAAGAAAAAGATATACAGCAATCGTATTAATTAATAT-----AATATCTTTT 2004
 Db 530 Asp-----GlnGlnSerValPheProAsnAlaThrAspAsnThrLeuIle 544
 2005 GAAGTATTTTAAAGTTATGATTAACCTGCAAGAAAGAGCAAAATGGAAGAA 2064
 Db 545 Thr-----LysLeuHisSerHisPheSerLysLysAsnAlaLysTyrGln 560
 2065 CTTATGAAATATTAATAAGAAAAAATGAGTTTCCAT----- 2106
 Db 561 -----ProArgPheSerLysThrGlnPheGlyValThrHisTyrAlaGln 576
 2107 -----TTGAAATTAATAGGACTATTTAGACAAATCAATA 2142
 Db 577 ValMetTyrGlnIleGlnAspTyrPheGlnLysAsnLysAspProLeuGlnAspLeu 596
 2143 GAACCTGTTAGATCACTTAAGAAACCTGCAAGATGATTAAGACAAATTAACAAC 2202
 Db 597 GluLeuCysPheLysAspSerSerAspAsnValValThrLysLeuPheAsnProAsn 616
 2203 GAAGCATGTGAACATCCCATATGCA----- 2229
 Db 617 IleAlaSerArgAlaLysLysLysLysAlaAsnPheIleThrValAlaAlaGlnTyrLysGln 636
 2230 -----ACAAACACCGG-----TGCTT 2247
 Db 637 GlnLeuAlaSerLeuMetAlaThrLeuGlnLysThrThrAsnProHisPheValArgLysIle 656
 2248 AAACCTCGGAGGACGACCAACCCCAATTAATAAAGAAATA----- 2292
 Db 657 IleProAsnAsnLysGlnLeuProAlaLysLeuGlnAspLysValValLeuAspGlnLeu 676
 2292 ----- 2292
 Db 677 ArgCysAsnGlyValLeuGlnLysIleArgIleThrArgLysGlyPheProAsnArgIle 696
 2293 -----GCACATACTTTAAAG----- 2310
 Db 697 IleTyrAlaAspPheValLysArgTyrTyrLeuLeuAlaProAsnValProArgAspAla 716
 2310 ----- 2310
 Db 717 GluAspSerGlnLysAlaThrAspAlaValLeuLysHisLeuAsnIleAspProGln 736
 2311 -----AGTCATACGAGAA 2325
 Db 737 TyrArgPheGlyIleThrLysIlePhePheArgAlaGlnGlnLeuAlaArgIleGln 756
 2326 GCACGAATCGGCTCTCATTAATTTGAAGAGGACACGACGAAGGTATATTAACGT 2385
 Db 757 AlaArgGlnGlnArgIleSerGlnIleIle---LysAlaIleGlnAlaAlaThrArgGly 775
 2386 GGGGTAGAGAAAGACTCAAGACAAT-----TTATGATTAATG 2430
 Db 776 TrpIleAlaIleGlyValTyrLysGlnAlaArgGlnHisIleThrValAlaAlaArgIleIle 795
 2431 ATAAACATTTTAATCTGATCTGTTTTCAAATGACCATGTGATGCAAGACACA 2490
 Db 796 GlnGlnAsnLeuArgAlaTyrIleAspPheLysSerTyrPro----- 809
 2491 GGTATGTATTAACAACAAGATTGTCTAGGAAGTGAATG-----GAA 2535
 Db 810 -----TyrTrpLysLeuPheSerLys 816
 2536 GTGATCGGGAACACATGCGTAAGATCAACGAGATGTTATTAATCCCTCAAGACAGA 2595
 Db 817 AlaArgProLeuLeuLysArgArgAsnPheGlnLysGlnLysGlnLysGlnLysGln 836

QY 2596 CATATA---TGTACATCCAAATTGGCAACATTACAAAGCATGATCACCACCTTAATGTT 2652
Db 837 ILeuGluLeuLeuSerAsnLeuThrAspSerThrGlnLysAspLysLeuGlnLys 856
QY 2653 AATATCTGTATGATTAGTTAATTAATTCCTTTTGGGGATGCTGCTTCTATCAGCAAAA 2712
Db 857 SerLeuLysAsp-----ThrGluSerAsnValLeuAspLeuGlnArgGlnLeuLys 873
QY 2713 TATGAGCAAAACAAGATATACAGATGTATTAAGAAAAGATATACCTTAAGGGCCCCAAA 2772
Db 874 ALAGluLysGluThrLeuLysAlaMetLysAspSerLysAspAlaLeuGlnAlaGlnLys 893
QY 2773 -----GAACTAACTGACCCCAAAACACCAG 2796
Db 894 ArgGluLeuGluLeuArgValGluAspMetGluSerGluLeuAspGluLysLeuAla 913
QY 2797 ACAACATTCGTGCTGAGCTATACCTTACGATTTTGCATATAGTGTATATATTCGAGCA 2856
Db 914 LeuGluLysLeuGlnAsnGlnLysArgSerValGluGlu-----LysVal 928
QY 2857 AGAGATCTCTGGCAAAAGCAAGGTGACATGCTAAAGCTGCAAGACATTTGGAACCTGT 2916
Db 929 ArgAspLeuGluGlnLysLeuGlnGluGlnLysLeuArgAsnThrLeuGlnLysLeu 948
QY 2917 TTTGGCTATATACATAAGTCACTCAA-----GGCAAGCAAAATGATTAATATATGAT 2970
Db 949 LysLysLysThrGluGlnGluGlnLysLeuGlnLysMetLysArgValAsnAspGluLysSer 968
QY 2971 GATCCCCCAAAATTTAAATTTAGAGCAAAATGTGTGGAGAGCTATATACGCAAAAGTA 3030
Db 969 ThrLysSerArgLeuGlnLysLysAspLysLeu-----GlnLysGluVal 984
QY 3031 TGGCAACCCATGAAATGTATTAATATTTGAAGATTAATCGGGACCAATCAACA 3090
Db 985 GlnGluLeuThrGlnLysLeuSerPheSerGlnLysLysAspLysGluValLeuGlnLysThr 1004
QY 3091 CAAAGTGTATTTCCGATATAGTATCATACACCATTTGATTTATTCGCCCAAAAA 3150
Db 1005 ArgValArgLeu-----GlnSerGluLeuAspAspLeu-----Thr 1016
QY 3151 TTAAGATGATGATGACCAATGGCAGATGTACTGCAAGGTGCAAGAAAGAGATGATGAT 3210
Db 1017 ValArgLeuAspSerGluThrLysAspLysSerGluLeuArgGlnLysLys----- 1034
QY 3211 AAGTTGAAGAGAAGTAAAGAGAGTGTAGGATTAAGATTAAGTGTCAAGGCTGTACGAAA 3270
Db 1035 LysLeuGlnGluGlnLysLeuLysGlnValGlnGlu-----AlaLeu 1047
QY 3271 GAGAGTGTACAGGTTTACGAAGTGCACAGAACCTTGTATGAATATATGATATATA 3330
Db 1048 AlaAlaGluThrLysAlaLysLeuAlaGlnGlnLysAlaLysLysLys----- 1063
QY 3331 GGATTATGGAAGAACAAGAT 3390
Db 1064 -----LeuGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 1074
QY 3391 GCACAAATGCTGCTT---AGTATAGTGTATGATGACCTTCCAGTACTGCCAAATATCAT 3447
Db 1075 PheAsnSerGlnValThrAlaArgSerAsnValGlnLysSer-----LysLysThr 1091
QY 3448 ATAGACAGCAATGTATATGATTTTTCGGAATTTATATCAACAACAATGTGTGCAAAAGT 3507
Db 1092 LeuGlnLysGlnLysValAlaValAsnAsnGlnLysAspGlnGlu-----LysLys 1108
QY 3508 AATAAAGTGTACTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3567
Db 1109 AsnArgAspAlaLeuGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 1126
QY 3568 GTTGAGACATATCTCCATGATACAGAAATTTGATGATGATGATGATGATGATGATGATGAT 3627
Db 1127 MetLysAspGlnLysGlnLysThrGlnGly-----GlnLysLysSerLeu 1141

QY 3628 TGTGATCAAAAAAGTCATGCTTAAGGATTAACGAAAAATATCCCTTTAGAGATTAACACAG 3687
Db 1142 TyrAspLeuLysValLys-----GlnGluSerAspMetGlnAlaLeuArgAsnGlnLys 1160
QY 3688 GACCATGATGTGCTGTGCTGTGAAGATGCAACCCGCA----- 3732
Db 1161 GlnLeuGlnSerThrLysAlaLysLeuGlnLysLysLysSerThrLeuGlnGlnVal 1180
QY 3733 ---AGGTTACAG-----ATAAACGAAA 3753
Db 1181 AlaArgLeuGlnGlnLysLeuGlnAlaGlnGlnLysSerAsnValGlnLysGln 1200
QY 3754 AAAAAAGCAAGAAAAAGATACGGAATGTAAACAGTAAATGATATCTTAAGAAAAAC 3813
Db 1201 LysLysLysValGlnLeuAspLeuGlnAspLysSerLysLys-----LeuAlaGlnGln 1218
QY 3814 GATGGAAGAAACAGATGAGAAATGTGCATCCAAAGAAATAGTATGATATCCCGAT 3873
Db 1219 ThrAlaLysGlnAlaLysAspLysLeuLysLysLys----- 1231
QY 3874 TGGCAATGCGGAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3933
Db 1231 ----- 1231
QY 3934 CAAAGTATATCGTACATTTCTTGGCAATGATATGATTAATTAATTAATTAATTAATTA 3993
Db 1232 -----LeuGlnGlnLysLeuSerGlnValGlnThrGln 1242
QY 3994 GTTAATTTAAAGAGCTTTCATCAATTCGACAGCAACAAACATTTCTCTCATGTGAT 4053
Db 1243 -----LeuSerGlnAlaAsnAsnLysAsnValAsnSerAspSer----- 1255
QY 4054 TATATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4110
Db 1256 ThrAsnLysLysLeuGlnLysThrSerPheAsnAsnLeuLysLeuGlnLysAlaGlnGln 1275
QY 4111 AAAATTCCTCCCGATTTTGAAGTCCATGTCTTACACATTTGAGATTTATGAGATTTT 4170
Db 1276 LysAlaLysGlnAlaLeuGlnLys----- 1283
QY 4171 TTATTTGAAACAGATATATCAAAAGTCAATGATGAGGAAATTAATTAAGACAAATA 4230
Db 1284 -----LysArgLeuGlnLysLeuGlnLysLeuGlnLysLysLysLysLysLys 1296
QY 4231 GATTCCTTTTCAAAAAATGTGTACCAAAAAATCTCTTAATGGAAGAACGCCAAGATGG 4290
Db 1297 GlnGlnLysGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1316
QY 4291 TGGCAACACATGATCATGAGATATGGGAACCTATGCTATGTGCATGATTAATTAATTCGG 4350
Db 1317 GlnLysGln-----ValSer 1321
QY 4351 GCAAAAAAGATGATTTTACGAAACCTAGCGTTACAAACGCAATTAATTAATTAATTAAT 4410
Db 1322 GlnLeuLysAspGlnLysGlnGln-----GluValAlaSerLys 1334
QY 4411 AGCACACTTTGGAGAAATTTGCCAAAGCACCCAGTTTATGATGATGATGATGATGATGAT 4470
Db 1335 LysAlaValThrGlnAlaLysAsnLysLysLysLysLysLysLysLysLysLysLysLys 1354
QY 4471 TACAGCAGCTATTTCTTACACGCAACAAATATTTGAAGATGTGCGAGGAAAAATGTAAG 4530
Db 1355 TyrAlaAspValValSerSerArgAspLysSerValGlnGlnLysLysLysLysLysLys 1374
QY 4531 TCAATGACCAAAATGAGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 4590
Db 1375 LysAsnGlnGlnLysLeuArgSerThrAlaGlnGlnAlaGlnGlnLysLeuAspArgAlaGln 1394
QY 4591 TATATGAAAAAAGAGAGTGTATTCACAGAT-----AAATATTAACAGATGAA 4644
Db 1395 ArgSerLysLysAlaGlnLysPheAspLeuGlnLysAlaValLysAsnLeuGlnGlnLys 1414
QY 4645 CGGCACAAAAAAGATTCGATAGACAAACATTTGCTATGATGATGATGATGATGATGATGAT 4704

Db 1415 ThrAlaLysValLysAlaGluLysValGluThrAspTyrArgSer 1434
OY 4705 AGC-----AATCCACAGATTACTTGACAGAAATTTACTGCTAGTTGT 4749
Db 1435 ThrLysSerGluLeuAspAlaLysValSerSerGluGlnTyr----- 1450
OY 4750 GGTGATAACTGGAAGCCCTCTGCTGACAAAGAAATATCAATTTGTGAA---AA 4806
Db 1451 -----ValGlnIleLysArgLeuAspGluGluLeuSerGluLeuArg 1464
OY 4807 CAGGCTACTATGATGCGCAAAACATTGTGGTGCACAAATTTATGAAATGACGAC 4866
Db 1465 SerValLeuGluGluLeuAspGluArgCysAsnSerAla-----IleLysAla 1480
OY 4867 AATATACTACATTTGCGATAAGATTAAGTCCAAAGATTAGTAAAGGACCAACACA 4926
Db 1481 LysLysThrAlaGluSerAlaLeuGluSerLysAspGluIleAspAlaAlaAsn 1500
OY 4927 GGTGCTTTAGTGGCAAAACAAAGCTCTAATTAATCACTAATCACTGAAATGACT 4986
Db 1501 AlaLysAlaLysAlaGluLysSerLysGluLeuGluValArgValAlaGluLeuGlu 1520
OY 4987 GAAGATGCTGCTTTCTCTCTGCTGCTACGTATATGTTT----- 5028
Db 1521 GluSerLeuGluLysSerGlyThrValAsnValGluPheIleArgLysLysAla 1540
OY 5029 -----CATGATTTGATGCGCAATTAACAGATCCAGAACTGAA----- 5067
Db 1541 GluIleAspAspLeuArgAlaArgLeuAspArgGluThrGluSerArgLysSerAsp 1560
OY 5068 ---GATGAAATGGTTGCGAAAGATGATGAGTGGCGCAACGAAAGGTACAAAT 5124
Db 1561 GluAspLysLysAsnThrArgLysGlnPheAlaAspLeuGluAla----- 1575
OY 5125 TTGGGTCAATACTCAAAAGAAAAAGAAAAAGAAAAATATAAAAGTCGATGGCGAC 5184
Db 1576 -----LysValGluGluAlaGluArgGluValValThrIleAspArgLeu 1590
OY 5185 AATATTTCTTATGAGTCCGCCCTGTGATGATGAAATATGTTTATGATTAGCA 5244
Db 1591 LysLysLysLeu-----SerAspIleIle 1599
OY 5245 GATATTAATCTAGTATGATTAATTGGAAATGAAAAACAAAGACCGAGAAATTTG 5304
Db 1600 AspLeuSerThrGluLeuAspThrGluThrLysSerArgLysIleGluLysSerLys 1619
OY 5305 AAGAAATATTTTACAAAAATGGAACATCACTGGCAAGAGAGTACTAGTACAGCA 5364
Db 1620 LysLysLeu-----GluGlnThrLeuAla 1627
OY 5365 AATCCCGTAGTACTGCGCGAAATTTTCTGGAACGAAATAGAAATGTGTGTGAC 5424
Db 1628 Glu----- 1628
OY 5425 GCAATGATTCGGGTACAAACGTGTAGGATGATGAAATAGTGAATAGTCAAGA 5484
Db 1629 -----ArgArgAlaIleGluGluLysSerLysAla 1639
OY 5485 AGTATGATGATTAATAAAATGTTCTGTCTACTCTCAGATGATGATTAATCTATG 5544
Db 1640 AlaAspGluGluIleArgLys-----GluValTyr 1649
OY 5545 AAAATCCGATGAAGTACTGCTATCAGTTTCTTGATGCTTTCCGGAATGGGGTGA 5604
Db 1650 GlnGluValAspGlu-----LeuArgAlaGluLeuAspSerGluArg 1663
OY 5605 GATTTTTCAAACTAAAGAAAGAAATGGAATGGAATGAGAGGGCGGTAAATGATTAT 5664
Db 1664 AlaAlaLeuAsnAlaSerGluLysLysIleLysSerLeuVal----- 1677
OY 5665 ACTTGTGTGATGATGAATGAATGAAGAAATGTCAGATGCTGTACCAATATATAA 5724

Db 1678 -----AlaGluValAspGluValLysGluGluLeuGluLysGluIleLeuAlaLysAsp 1695
OY 5725 AATTTTATGATGATGAGAACCAACAGTATGAAACAAATCAAAATATGCTGCAAT 5784
Db 1696 LysLeuValLysAla---LysArgAlaLeuGluValGluLeuGlu-----GluVal 1711
OY 5785 AAGACAAATATATTTCCGACATCTGTGGCAAAAGATCAGAGACGCTCGCAATAT 5844
Db 1712 ArgAspGluLeuGluGluGluLysAspSerArgSerGluLeuLysSerLysArgArg 1731
OY 5845 TTACACAAACATTAATAAAATTTGCAAAATTAAGTGAGATGCTGCAATTAAGTGT 5904
Db 1732 LeuThrThrGluValGluAspIleLysLysTyrAspAlaGluValGluGlnAsnThr 1751
OY 5905 ATGAAAGATGTGTCACACAGCATTAATCACTGATGTAATGTCAAATATGCCGATCA 5964
Db 1752 LysLeuAspGluAlaLysLysLysLeuThrAsp---AspAlaAspThrLeuLysGln 1770
OY 5965 TTGACGATGAAACCAAGAAAGTTGAAAGAAAGTGAATTTGTCAAGTCCACAGTCCA 6024
Db 1771 LeuGluLysGluLysLys----- 1777
OY 6025 CCACGTGACGAAGGAACACCGTCAACGCGTATCAGTATGCAAAAGCGACGCA 6084
Db 1778 -----LeuAsnGluSerGluArg 1783
OY 6085 TCGAAAAAGAGCGAAACACCGCCCTTACAAACACCGGAAAGTGGAAATCTA 6144
Db 1784 AlaLysArgLeuGluSer-----GluAsnGluAspPhe 1795
OY 6145 ACAACGAATGCGAGCAACCAACGACGACGACGACACAAACGAAACGA 6204
Db 1796 LeuAlaLysLeuAspAlaGluValLysAsnArgSerArgAlaGluLysAspArgLys 1815
OY 6205 -----ACATCAACGACAAACACA 6225
Db 1816 TyrGluLysAspLeuLysAspThrLysTyrLysLeuAsnAspGluAlaIleThrLysThr 1835
OY 6226 GAATCTGACGTGGCACAATGTTAAAGCCATCTCTTGATTAACCGATGACGAGGCT 6285
Db 1836 GlnThrGluIleGlyAla-----AlaLysLeuGluAspGlnIleAspGlu----- 1850
OY 6286 GGAATAGAGGTTGTATCCAAACAAAGTATGACAAATATCTTAATGGGTTGATTGTA 6345
Db 1851 -----LeuArg 1852
OY 6346 GGTAACTTAAGAAATGAAATGCGATATGCTCTAGGAGAAATATATAGT 6405
Db 1853 SerLysLeuGluGluGluGluAlaLysAlaThrGlnAlaAspLysSerLysThrLeu 1872
OY 6406 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6465
Db 1873 GluGluGluIleAspAsnLeuArgAlaGlnIleGluAsp-----GluGlyLysIleLys 1890
OY 6466 GAGGCTTTATTAATA---TGTCACGCAATGAACTCAATTTTGTGGTAAATATATA 6522
Db 1891 MetArgLeuGluLysGluLysArgAlaLeuGluGluGluGluLeuArgGluThr 1910
OY 6523 ATGAAATATCCGACAGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAAT 6582
Db 1911 ValGluGlu-----AlaGluAspSerLysSerGluAla-----GluGlnSerLys 1925
OY 6583 AGAATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6642
Db 1926 ArgLeuValGluLeuGluLeuGluLysAlaArg-----ArgAsnLeuGln 1940
OY 6643 AATGATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6702
Db 1941 LysGluIleAspAlaLysGluIleAlaGluAspAlaLysSerAsnLeuGlnArgGluIle 1960
OY 6703 AAGAAAAACAGATTAATAAAAGATGAAGATTAATTAATTAATTAATTAATTAATTAAT 6762
Db 1961 ValGluAlaLysGluLysArgLeuGluGluSerIleAlaArgThrAsnSerArgSer 1980

QY 6763 AAAAATTATTTGGAGAGAAATGATATGATTAATCTTACAGAGAAAC 6822
 Db 1981 ArgysArgLeu---GluAlaGluIleAspAlaLeuThrIleAlaVal---AspAlaGlu 1998
 QY 6823 GAAAAAAGAAAAATTAAGTAATTAACAGTACATGACATGACAAACAGCCCTCC 6882
 Db 1999 GluLysAlaLys-----AsnGlnGlnIleLysGluLysLysIleGluThrGlu 2015
 QY 6883 CTTGAGAGCTTTGTAAGAGCCCAATTTTTCAGATGTTCCAGAGATGGCAGAAAG 6942
 Db 2016 LeuLysGluThrArgLysLys-----PheGlyLysLeu 2026
 QY 6943 TTTTGTATTAAGAGAGAACTGTTA-----AAATGAGAGCCGCTGTAGAA 6996
 Db 2027 GluLysThrLysThrLysGluPheLeuValValGluLysLeuGluThrAspLysArg 2046
 QY 6997 TATGAGCTGATAGTGTATTAAGAGAGTATGACACAGAAATGTCAGAGCCGTGTACA 7056
 Db 2047 AlalysLysGluAlaAlaAspGluGlnGlnGlnIle---LeuThr 2060
 QY 7057 TATCAAAATTTTATTAAGAGAGTATGATATGAAAGACAAAGAAAGTTCAA 7116
 Db 2061 ValGluLysAspLeuArgLysLysLeuSerGluIle-----SerLeuLeu 2075
 QY 7117 AAGGATTAAGATGCAAAAGATTAAGATTAATTCCTTCTACTGAAAGACATAGAGAG 7176
 Db 2076 LysAspAlaIleAspLysLeuGlnArgAspHisAspLysThrLysArgGluLeuGluThr 2095
 QY 7177 GCACATGCTGCTCATGATATTTTAACATGAATTAAGAAATTAATTCGCG 7227
 Db 2096 GluThrAlaSerLysIleGluMetGlnArgLysMetAlaAspPheGly 2112

RESULT 12
 NDM1_YEAST STANDARD; PRT; 2748 AA.
 ID NDM1_YEAST STANDARD; PRT; 2748 AA.
 AC C00402;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Nuclear migration protein NDM1.
 OS NDM1 OR YDR150W.
 GN Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae;
 OC NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=ATCC 28383 / FL100;
 RA MEDLINE=9207907; PubMed=1745235;
 RA Kormanec J., Schaffl-Gerstenschlaeger I., Zimmermann F.K.,
 RA Perecko D., Kuentzel H.;
 RA "Nuclear migration in Saccharomyces cerevisiae is controlled by the
 RT highly repetitive 313 kDa NDM1 protein.";
 RT Mol. Gen. Genet. 230:277-287(1991).
 RU 1- FUNCTION: CONTROLS NUCLEAR MIGRATION. NDM1 SPECIFICALLY CONTROLS
 CC THE INTERACTION OF THE BOD NECK CYTOSKELETON WITH THE PRE-
 CC DIVISIONAL G2 NUCLEUS PERHAPS BY RECOGNIZING G2-SPECIFIC
 CC CYTOLASMIC MICROTUBULI OR OTHER COMPONENTS OF THE NUCLEAR
 CC ENVELOPE.
 CC 1- MISCELLANEOUS: ADDITIONAL REGIONS OF LOWER HOMOLOGY TO THE REPEAT
 CC CONSENSUS (ALWAYS STARTING WITH PROLINE) ARE FOUND IN BOTH
 CC FLANKING DOMAINS OF THE TANDDEM REPEATS.
 CC 1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 IS THE INITIATOR.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: X61236; CAA3554.1; -.
 DR PIR: S19052; S19052.
 DR SGD: S0002557; NDM1.
 DR InterPro: IPR001849; PH.
 DR Pfam: PF00169; PH; 1.
 DR SMART: SM00233; PH; 1.
 DR PROSITE: PS50003; PH; 1.
 DR PROSITE: PS50003; PH; 1.
 KW Repeat.
 FT DOMAIN 593 1384 12.5 X TANDDEM REPEATS.
 FT REPEAT 593 656 1.
 FT REPEAT 657 727 2.
 FT REPEAT 728 798 3.
 FT REPEAT 799 862 4.
 FT REPEAT 863 926 5.
 FT REPEAT 927 990 6.
 FT REPEAT 991 1054 7.
 FT REPEAT 1055 1118 8.
 FT REPEAT 1119 1182 9.
 FT REPEAT 1183 1246 10.
 FT REPEAT 1247 1310 11.
 FT REPEAT 1311 1374 12.
 FT REPEAT 1375 1384 13 (INCOMPLETE).
 FT REPEAT 1375 1384 13 (INCOMPLETE).
 FT DOMAIN 2573 2683 PH.
 SQ SEQUENCE 2748 AA; 313202 MW; B2FFB67C9F6211AE CRC64;
 Alignment Scores:
 Pred. No.: 3, 07e-08 Length: 2748
 Score: 322.50 Matches: 610
 Percent Similarity: 31.75% Conservative: 437
 Best Local Similarity: 18.50% Mismatches: 1112
 Query Match: 1.66% Indels: 1140
 DB: 1 Gaps: 164

US-10-087-013-1 (1-10628) x NDM1_YEAST (1-2748)
 QY 1495 CACAAATCAGATTAATGATCTGAA----- 1518
 Db 7 HisLysLysAsnAsnAspLysAspSerSerAlaGlyGlnIleAlaAsnSerIleAspAsn 26
 QY 1519 CGTGTAAATATTAAGAGACTTAATTAACCTCCATCGGGGTGAAGCCTACTAATATACATGTC 1578
 Db 27 SerLeuSerGlnGluSerValSerThrAsnGlyAlaThrArgMetAlaAsnLeuLysAla 46
 QY 1579 CTTATAGTGTAAAT---GAACAAGTGATATTAACAAAA----- 1617
 Db 47 AspLysGluSerGlyAspGluGlyAspLysThrLysArgPheSerIleSerSerIle 66
 QY 1618 -----TTAGAAAAATTTTGTAAACAGCTCACT----- 1644
 Db 67 LeuSerLysArgGluThrLysAspValLeuProGluPheAlaGlySerSerIleAsn 86
 QY 1645 -----AATTCAAGATTAAT-----AAT 1665
 Db 87 GlyValLeuThrAlaAsnSerSerLysAspMetAsnThrLeuGluLeuSerGluAsn 106
 QY 1666 CAAAAATGCGATCTTATTAAGATGAAGAAATTAATAGATTAAGTAACTGAACAATAAT 1725
 Db 107 LeuLeuValGluLysArg-----LysLeuGlnSerSer 117
 QY 1726 ACTGAATCAATTAATGATTAATCTTAAGATTAATCAATTCATTAATTTTGAATATGG 1785
 Db 118 AsnGluLysLysAsnGluGln----- 124
 QY 1786 GTTACATATTTATTAAGGATTAATTAAGTGAAGTGAAGTAACTTATTAAT 1845
 Db 125 -----IleLysSerLeuLysGlnIleLysGluSerLeuSer 136
 QY 1846 AATCAACACACGATGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1905
 Db 137 Asp----- 137
 QY 1906 GTTAAACAAAAAGAGAGAGATGAATGATTAAGAACTGTTCAACAAAA----- 1959

```
Db 138 -----LysIleGluGluLeuThrAsnGlnLysSerPheMetLysGluLeuAsp 154
OY 1960 -----AAGATATATACGCAATCCTATTATAGTATATATATATATATCTTTTGAAGT 2010
Db 155 SerThrLysAspLeuAsnTPaspLeuGluSerLysLeuThrAsnLeuSerMetGluCys 174
OY 2011 TATTTTATTAAGTTATGATTAACCTTGACAAAGATGACCAAAATGGAAGACTTATG 2070
Db 175 ArgGlnLeuLysGluLeuLysLys-----LysThrGlnLysSerThrPasnAspGluLys 192
OY 2071 GAAATATATAAGAAAAAATGACTTTTCCAAATTGGAATATATAGGACTATTTA 2130
Db 193 GluSerLeuLysLeuLeuLysThrAspLeuGluIleLeuThrLeuThrLysAsnGlyMet 212
OY 2131 GAGATGCAATAGAACTCTGTAGATTAAGAAACTGCCAGATATGATAAGAC 2190
Db 213 GluAsnAspLeuSerSerGlnLysLeuHisTyrAspLysGluIleSerGluLeuLysGlu 232
OY 2191 -----AATATACAAAGCAAGCATGTGAACATCCCATTAATGCAACACAA 2235
Db 233 ArgIleLeuAspLeuAsnAsnGlnAsnAspArgLeuLeuIleSerValSerAspLeuThr 252
OY 2236 AACCCGTGTCTTAACCTGCTGAGGACGACACCCACTAAATATATAAGAAATAGCA 2295
Db 253 SerGluIleAsnSerLeuGln-----SerAsnArgThrGluArgIleLysIleGlnLys 270
OY 2296 CAATACTTAAAGAGATGCAATCGAGAGAACGCAAAATCGTGTTCATTAATTAAGAA 2355
Db 271 Gln-----LeuAspAspAlaLys--AlaSerIleSerSerLeuLys 283
OY 2356 GGAAGGACACAGCAAGTATATATAACGTGGGGGTAGAGAAAGCACTTCAAGCAAT 2415
Db 284 ArgLysValGlnLysLysTyrGln----- 292
OY 2416 TATGTAGATATATGATTAACATTCATCTGATCTGTGTTTCAATGAGCACATGT 2475
Db 293 -----LysGlnHisThrSerAspThrThrValThrSerAspPro--- 305
OY 2476 GATGCAAGGACACAGATGTGTATACAAACAGATTTGTCTGAGAACTGAATGGGA 2535
Db 306 AspSerGluGlyThrThrSer----- 312
OY 2536 GTGATCCGGAACATCGGTAAGATCAGAAAGATGTTATATGCTCTAGAAAGACA 2595
Db 313 -----GluGlnAspIlePheAspIleValIle----- 321
OY 2596 CATATATGATCCAAATTTGGAACATTTACAAACGATGATCACCCACTTAATGTAAT 2655
Db 322 -----GluIleAspHisMet--IleGluThrGlyProSerValGluAsp 335
OY 2656 ATTGTGATGATTAGTTAATAATTCCTTTTGGGGGATGTTCTTATCAGCAAAATAT 2715
Db 336 IleSerGluAspLeuValLys-----LysTyr 344
OY 2716 ---GAAGCAACAAAGATATACGATGTATAAGAA-----AAGATTAACCTAAAGGC 2766
Db 345 SerGlnLysAsnAsnMetIleLeuLeuSerAsnAspSerTyrLysAsnLeuLeuLys 364
OY 2767 CCCAAAGAAAGTACTGACCCAAACACAGACAACTATCTGCGACTATACGTATACGT 2826
Db 365 SerGlnSerIleAspLysProLysAspArgLysLeuMetThrLysGluVal----- 381
OY 2827 TTTCGATATATAGTATATATTCGAGAGAGATCTCTGGGAAAGAAAGCGTGACATG 2886
Db 382 -----AlaGluAsnLeuAsnMet 387
OY 2887 GTAAGGTGCAAGACATTTGAAACTGTTTGTGTAATATACATTAAGTCACTCAAGGC 2946
Db 388 IleAlaLeuPro----- 391
OY 2947 AAAGCAATGATTAATATGATGATGCCCCCAATATTTAAATGAGGAGAAATTTGG 3006
----- 391

Db 392 -----AsnAspAsp----- 394
OY 3007 TGGGAGCTAATATAGACCAAGATATGGAAACCAATGATATTAATATTTGAAG 3066
Db 395 -----AsnTyrSerLysLysGluPheSerLeuGlnSerHisIleLysTyrLeu--- 410
OY 3067 GATTAATCGGACACCA-----TCACACAAAGATTAATTCGGCATATGAT 3117
Db 411 GluAlaSerGlyTyrLysValLeuProLeuGluGluPheGluAsnLeuAsnGluSerLeu 430
OY 3118 CATACCCCTTGATGATATATCCACAAAATTAAGATGATGACCAATGGCGACAA 3177
Db 431 SerAsnProSerTyrAsnTyrLeuLysGluLysLeuGlnAlaLeuLysLys----- 447
OY 3178 TGGTACTGCAAGTGCAGAAAAGAGTGTGTAAGTGAACAG----- 3222
Db 448 -----IleProIleAspIleSerThrPheAsnLeuLysGluProThrIleAspPhe 465
OY 3223 -----AAGTGAAGAGGTGAAGATAGGATATATGCT--- 3255
Db 466 LeuLeuProLeuThrSerLysIleAspCysLeuIleIleProThrLysAspTyrAsnAsp 485
OY 3256 -----CAAGCTGTACAGAAAGAGGTGACAGGTGTACAGAACTGCAAGACTTGT 3309
Db 486 LeuPheGluSerValLysAsnProSerIleGluGlnMetLysLysCysLeuGlnAlaLys 505
OY 3310 AATGAATATATATATATATATGATATGAAAGAACATGCAAT----- 3354
Db 506 AsnAspLeuGlnSerAsnIleCysLysThrLeuGlnGluArgAsnGlyCysLysThrLeu 525
OY 3355 -----ATAATA 3360
Db 526 SerAsnAspLeuTyrPheSerMetValAsnLysIleGluThrProSerLysGlnTyrLeu 545
OY 3361 TCAGTAAATACAAAGATTTACATGACACACAAATGTGTGTAATATAGTAT 3420
Db 546 SerAspLysAlaLysGluTyrAsp-----GluValLeuIleAspThrLysAlaLeu 562
OY 3421 GAAGCTTCAGTACTGCCAAATATCATATGACAGAAATGTTATGAAATTTTTCGGA 3480
Db 563 GluGly-----LeuLysAsnProThrIleAspPheLeuArgGlu 575
OY 3481 TTATACCAAAATGCTGCAAAAGTAAATGAAGTGTACTAGTAT----- 3528
Db 576 -----LysAlaSerIleAspTyrLeuLeuLeu 585
OY 3529 ---GAAGTGTCTGCTATGTTGATTAACACACAGCTAT-----GAAATGTGGAGCA--- 3576
Db 586 LysLysGluAspTyrValSerProSerLeuGluIuThrLeuValGluHisAlaLysAlaThr 605
OY 3577 ---TATCTCATATGATACAGAAATTTTGTGAT-----TGTCAAGTCA--- 3615
Db 606 AsnHisIleLeuSerAspSerAlaTyrGluAspLeuValLysCysLysGluAsnPro 625
OY 3616 CAATATGACTTTTGTGATGAAAAAGT-----GATGTGAAGATTAACGAA 3660
Db 626 AspMetGlnPheLeuLysGluLysSerAlaLysLeuGlnHisThrValIleSerAsnGlu 645
OY 3661 AAATATGCTTTAGAGTAAACACAGACCAATGATGTGCGTGTGTTAAAGTGA 3720
Db 646 AlaTyrSerGluLeuGluLysLysLeuGluGlnProSerLeuGluIuThrLeuValGluHis 665
OY 3721 TCGAAACCGCAAGGATACAGATTAATAACGAAATAAAGCGAGAAAGATAGCGAA 3780
Db 666 AlaLysAlaThrAsnHisIleLeu-----LeuSerAspSerAlaTyrGluAspLeuValLys 684
OY 3781 TGTAAACAGTGAATGATATA-----CTTAAAGAAACATGGAAGAAACAAAGTAGAA 3834
Db 685 CysLysGluAsnProAspMetGluPheLeuLysGluLysSerAlaLysLeuGly----- 702
OY 3835 GATTTGATCCCAAAAAGATATGATATCCGATATCCGATGCGCAATGCGGAATATTAAT 3894
Db 703 -----HisThrValIleSerAsnGluAlaTyrSerGluLeuGlnArgLysTyrSerGlu 720
```


Db 1255 HisAlaLysAlaThrAspHisHisLeuLeuSerAspSerAlaTyrGluAspLeuValLys 1274
QY 5902 TGTATGAAGATGTGTCTCACACACGATTAAGTATGATATAGTCAAAATATGCCCCGA 5961
Db 1275 CysLysGluAsnProAspMetGluPheLeu 1284
QY 5962 TCATTAAACGATGACCAAAAGAAAGTGAAGAAAGTGAATGTCAAGTCCACGAGGT 6021
Db 1285 -----LysGluLysSerAlaLysLeuGlnHisThrVal 1295
QY 6022 CCAACGCTGTACGAGGAGAAACACGCTCAACGCGGTATCAGTATTCGAAAGAGAGC 6081
Db 1296 -----ValSerAsnGluAla 1300
QY 6082 GCATCGAAAAAGAGCGAAACAGCCCGCTACAAACAGCCGAAAAATGCAAAAT 6141
Db 1301 TysSerGluLeuGluLysLysLeuGluGlnProSer 1315
QY 6142 CTAAACACAGAAATGCGAGGACAAACAGAACCCGAGACGACACACAAACAGAAA 6201
Db 1316 LeuValGluHisAlaLysAla 1322
QY 6202 CGACATCAACAGCAACACACAGATCTAGCTGGGACATGGTAAAGCCATTCTT 6261
Db 1323 ---ThrAsnHisLeuLeuSerAspSerAlaTyrGluAspLeuValLysCys 1339
QY 6262 TCCAATTAACACAGATGACGAGGCGTGAATGAGCGTTGTAATCCAAACGTAATGGACAA 6321
Db 1340 LysGluAsnProAsp 1352
QY 6322 TATCTTAATGGGGTGTATGTAGTAAAGTCAAAAGAAATGCAATATGTATG 6381
Db 1353 ---AlaLysLeuGlnLysHisThrValValSerAsnLysGluLysSerGlu 1367
QY 6382 CCTCCTAGAGAAAAAATTATGTATTAATATACATATTTAAATTATGAAACTGA 6441
Db 1368 ---LeuGluLysLysLeuGlnProSerLeuGluLysLeu 1380
QY 6442 AATAACCGTGAACATGATTAAGAGAGCGTTTATTAATGTGCGACCAATACAACTCA 6501
Db 1381 ValLysHisAlaGluGlnLeuGlnSerLysLysLeuSerLysSerAspPheAsnThr 1399
QY 6502 TTTTGTGGTAAATATATATTAATGAAATCCTGACAGAAATGAATGCAAAATGA 6561
Db 1400 -----LeuAlaAsnProSerMetGluAsp 1407
QY 6562 ACAATTCAGATGAATTAAGAAATATATGTATATCATATGCTGATTAAGATATG 6621
Db 1408 ---MetAlaSerLysLeuGlnLysLeuGlnLysLeu 1419
QY 6622 TTTTGTGAACATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6672
Db 1420 -----ValSerAsnAspGluLysLeuAlaLeuLysAsnThrLeuLys 1434
QY 6673 AGTGAACACCATCTCAATGAAATTAATTAAGAAAAAAGCAG-----GATTAATAAAAA 6726
Db 1435 ProAspValGluLeuLeuAspSerLysLeuLysGluLysHisLeuLysAspThrThr 1454
QY 6727 GATGAAGATTAATGTAATTAATTTGGAGAGAAAAATTAATTAATTTGGAGAGAAATG 6786
Db 1455 TyrAsnGluLeuValSerAsnPheAsnSerProThrLeuLysPheLeu-----GlnLys 1473
QY 6787 ATATATGATTAATCACTATCATCTACAGCAAGCAAGAAAAAGAAATTAAGATTAAT 6846
Db 1474 AlaLysSerLysGluLysArgLeuLeuLysProAsnGlu 1486
QY 6847 TACCACTCAATGACATACCAAACTG-----ACGCTTCCTGAGAGTTGTAAAA 6900
Db 1487 -----TyrLeuAspLeuAsnArgLysLeuAlaThrThrProSerLysGluGlnLys 1502
QY 6901 AGGCCCAATTTTGAATGTGTACAGAAAGGCGAGAAAGATTTTGTAAATTAAGAGAG 6960
Db 1503 -----AspAsnPheCysLys 1507
QY 6961 GAACTATGTTAAATTTGAGCGCGGCTGT-----AAGCAATATGAG 7002
Db 1508 -----GlnLeuGlnCysTyrAlaLeuAspSerLysGluLysArg 1521
QY 7003 TGTATGTAGTATGACGCGTAAACAGAAATGTCGAGAGCGGTGTAAATATCA 7062
Db 1522 LeuLysAsnSerLeuGlnAsnProSerLysLysPheLeuGlnAsnAlaLeuLeu 1541
QY 7063 AATTTATTAAGAAAGTGAAGAACTGAATTAAGAAAGCAAGAAAGTTCAAAAAGAT 7122
Db 1542 AspLeuValLeuValAspLysThrGluLysGlnAlaMet 1554
QY 7123 AAAAGTGGCAAAAAGTATGATATATCTCTTACTGAAGAGACATGACAGCAACA 7182
Db 1555 -----LysAspAsnAlaSerAsnLysSerLeuLysProSerThr 1568
QY 7183 TGTGCTCATGAATATTTAAACATGAATTAAGAAATTAATGTCGCAATAGAGTTGCT 7242
Db 1569 LysValLeuAspPheValThrMetProAlaProGlnLeuAlaSerAlaGlu 1585
QY 7243 TGTATGCAAAAACCTTTCTTACACACTACCAAAACACAAACAAATCATCCGAT 7302
Db 1586 -----LysSerSerLeuGlnLysArgThr 1596
QY 7303 GCTAATGATATGCGAAGATCGCTGATTAATGTCCTGAAGAAATTAAGAGTGCAGGT 7362
Db 1597 LeuGluAsnGluLeuLysAlaLeuGlnLysVal 1614
QY 7363 CCTGAACCTTCAAAAAGAGGATCTATGATTCATACAAAAAATTAAGTCAATTAATA 7422
Db 1615 ProAsnLeuGluLysPro 1620
QY 7423 CCTATGAATGTGTGAGAAAGCAGCATATTTATCTAAAGAGCAAGAAATTAATG 7482
Db 1621 -----IleValAspAsnAlaSerLysAsnAsp 1629
QY 7483 GATATTAACCTTGAGAAAAATTT-----ATACCTATGAGTGTACAAAGCAAAAGAA 7536
Db 1630 ValLeuAsnLeuCysSerLysPheSerLeuValProLeuSerThrGluGluLysAspAsn 1649
QY 7537 AGTAAATATGTTGACATAT-----ATATATCTTGCGATCTTAAGAACTTATGA 7590
Db 1650 MetArgLysGluHisThrLysLysLeuAsnLysLeuGlnLysProSerLysAspPheLeu 1669
QY 7591 CCGTAAATATATATGAGAAAGAAACCCCTGTAAGAAATAGAGAAAGAAATCGTTTAAG 7650
Db 1670 LysGluLys 1674
QY 7651 GTACATTTATGATGAAATGTTACAAAAATTCAAAGTTC-----TATCAGAGAAAAAGAA 7707
Db 1675 -----LysTyrGlnMetLeuLysLysSerLysHisAspLysTyrGluLysGlnGlu 1691
QY 7708 GTATGCTACCTCCAGAGAGACATATAGTCTTAAGAAAT 7749
Db 1692 AlaLeuGluAsnProGluLysThrPheLeuLeuGlnLysAlaSerAlaLeuGluLysGlu 1711
QY 7750 ---TTAGATGAATTAATTAATGAAAGACTTAAG-----GATAGTATTAATTCCTCA 7797
Db 1712 LeuValSerGluValGluLeuAspArgMetLysGlnMetLysPheSerProAspPheAsp 1731
QY 7798 AAAATGCTTGTGCAACTGCAACCAATGAAGCAATTAATTAATTAATTAATTAATTAAT 7848
Db 1732 TyrMetGlnLysAlaAlaArgAsnGluMetVal-----LeuLeuArgAsnGluLys 1750
QY 7849 -----TTCAATCAGAGAAAGGCTGC 7869
Db 1751 GluAlaLeuGlnLysLysLysGluLysProSerLeuThrPheLeuLeuGlnLysAlaAla 1770
QY 7870 GCAATGAATCAATA---TGTGATACTATGAATATAGTTCCGTGATC----- 7915
Db 1771 GlyMetAsnLysLysLeuValAspGlnLysGluLysThrAspGluThrIleArgLysCysAs 1790

QY 7916 -GGGTCACACTACTAGAG---GAACAGATGTTACGAATGGTGCTTACTTACCTCCC 7971
DB 1790 nHSPProthrinArgMetGluLeuGluSerCysHISHisLeuValLeuLeu-- 1809
QY 7972 GTACGAATATAATATAGAGTTTGTGATATACATATATAGGAAGAGGAATATAAAT 8031
DB 1810 ----AspGlnAspGluTyrSerThrLeuArgGluProLeuLysAsnArgAsnValGluAs 1828
QY 8032 AAGAGTAGAATATATACACAGATGTACAAAGCTTTCGTTGCTGGTGGAGTGTAAAT 8091
DB 1828 pleu-----LleasnThrLeuSerLysLeuAsnTyrLleAlaLleProAsnThrLleTy 1846
QY 8092 AGAAAAAGATATTTGGAAGACATGACGTGCAAAAGCACAGACATGCAAAATTTTACA 8151
DB 1846 rGlnAspLeuLleGlyLysTyrGluAsnProAsnPheAspTyrLeuLysAspSerLeuAs 1866
QY 8152 AAGGAAGATGATGATTTGAAGCCATATACATTAATCAAGATAGCTG----- 8203
DB 1866 nLysMetAspTyrValAlaLleSerArg---GlnAspTyrGluLeuMetValAlaLysTy 1885
QY 8204 -----GACATTAAGACAGATC-----CACCTGTGTAT 8229
DB 1885 rGluLysProGlnLeuAspTyrLeuLysLleSerSerGluLysLleAspHisLleValVa 1905
QY 8230 GATTATATACCTCAGCGTTTCGATGATGATGATGATGATGATGATGATGATGATGAT 8289
DB 1905 lPro-----LeuSerGluTyrAsnLeuMetVal----- 1914
QY 8290 CTGATGGAGAAATGCAAAAT-----TTAA 8316
DB 1915 -----ThrAsnTyrArgAsnProSerLeuSerTyrLeuLysGluLysAlaValLeuAs 1932
QY 8317 AATATCATGTATC----- 8329
DB 1932 nAsnHisLleLeuLleLysGluAspAspTyrLysAsnLleLeuAlaValSerGluHisPr 1952
QY 8330 -ACTGTAACACATCTGACAGATGCAAGATGATGATGATGATGATGATGATGATGATGAT 8383
DB 1952 oThrValLleHisLeuSerGluArgHisLeuLeuAsnLysValLeuValAspArgAs 1972
QY 8384 -----AGTGTAAACGATGTCAGATGATGATGATGATGATGATGATGATGATGATG 8430
DB 1972 nAspPheAlaThrMetSerThrSerLleGluLysProThrLleAspPheLeuSerThrLy 1992
QY 8430 ----- 8430
DB 1992 sAlaLeuSerMetGlyLysLleLeuValAsnGluSerThrHisLysArgAsnLysLysLe 2012
QY 8431 -----AAATCTGTATTCGA 8444
DB 2012 uLeuSerGluProAspSerGluPheLeuThrMetLysAlaLysGluGlnGluLleLle 2032
QY 8445 TATACATCAATCAATATACAAAGATTTGTATGAACA-----CC 8483
DB 2032 eLleSerGluLysGluTyrSerGluLeuArgAspGlnLleAspArgProAsnLeuAspVa 2052
QY 8484 AATATATACAAATCTCTACTATGATCAT-----GTCAAAATTTT---GTACAAA 8534
DB 2052 lLeuLysGluLysAlaLlePheAspSerLleLleValGluAsnLleGluLysArgGln 2072
QY 8535 GTTGAAAACCTTTTAAAGCAATGT-----TCTGTGACAGCTTTTCTGAATATCT 8585
DB 2072 nLeuValAsnThrThrSerProCysProLleThrTyGluAspLeuLysValTyrAl 2092
QY 8586 TCATGAA-----ACAAGTAGTGTGATTAATTAATTAAGAAATGATGTTCTTC 8639
DB 2092 aHisGlnPheGlyMetGluLeuLysLeuGln---LysProAsnLysLeuSerGluLysAla-- 2110
QY 8640 CAATATACGAACATATGCTTTCGAGAAACCAAAAGTTATTAAGAACCTTCAGAGTTG 8699
DB 2111 -----GluArgAlaGluArgLleAspGluGlnSerLleAsnThr 2123

QY 8700 TACACTACCTTTCAGAACATTCATGGATTAATTTCTACCGATCAAAACAAAGATGATG 8759
DB 2123 rThrSerSerAsnSerThrThrThrSerSerMetPheThrAspAlaLeuAspAsnLle 2143
QY 8760 TAAGGAATTCACAAACTTTTACCTTTCGTCGAGAAATGATATGATTAATTAATTCGATA 8819
DB 2143 eGluGluLeuAsnArgValGluLeuGlnAsnAsnGluAspTyrThrAspLleIleSerLy 2163
QY 8820 TTGGAAC-----GCATACCTTGTCTTAATGATTC 8849
DB 2163 sSerSerThrValLysAspAlaThrLlePheLleProLalaTyrGluAsnLleLysAsnSe 2183
QY 8850 AGATGATACAAAGGCTGA---TTGATTCCTCCAGACAGACATTTATGTCAAGACC 8906
DB 2183 rAlaGluLysLeuGlyTyrLysLeuValProPheGluLysSerAsnLleAsnLeuLysAs 2203
QY 8907 TATCAGTCATATTAATATGAAAGCGTATTAAGAAATTTTAAAAAAACTTCTTAC 8966
DB 2203 nLleGluAlaProLeuPheSerLysAspAsnAsp---Th 2216
QY 8967 TTCTGCTTCAGTCAAGACAAATGTTAGGTCAAAATATATATCGAGACAGATGTTG 9026
DB 2216 rSerValAlaSerSerLleAspLeu-----AspHisLeuSe 2228
QY 9027 CTTTGAGCGCAATGATATATGTTATGCATTTTCCGAT----- 9066
DB 2228 rArgLysAlaGluLysTyrGlyMetThrLeuLleSerAspGlnGluPheGluGluTyrH 2248
QY 9067 -ATATATTA-----GGACGATATGATGACACTTCAT 9101
DB 2248 sLleLeuLysAspAsnAlaValAsnLeuAsnGlyGlyMetGluGluMetAsnProLe 2268
QY 9102 ATCTGAAAAATTTAAAAATATTTGAAACATCAATGACAAACGAAATGTAAMAC 9161
DB 2268 uSerGluAsnGlnAsnLeuAlaLysThrThrAsnThrAlaGlnGluLys----- 2285
QY 9162 ATGCTGGAAAAATATATAGACGTACATATGCGACCTATGTATGTCGATATTAATTC 9221
DB 2286 -----AlaPheGlnAsnThr 2290
QY 9222 TACTCAAAATGTAACATTAATGATGAGATGCTGCAATTCACAAAGATGAAACCTAA 9281
DB 2290 rValProHisAsnAspMetAspAsnGluGluValGlyTyrGlyProAspAspProThr-- 2309
QY 9282 TCAGTTTCTTCGTTGTTAATGATGCGCAAGACAGCATGTAGAGAAAAGAACATGT 9341
DB 2310 -----PheThrValArgGlnLeuLysLysProAlaGlyAspArgAsnLeuL 2325
QY 9342 AAGTATTCATTTAAAAACAAATGTCCTGTCACAAAGATATATTT----- 9390
DB 2325 eLeuThrSer---ArgGluLysThrLeuLeuSerArgAspAsnLleMetSerGlnAs 2344
QY 9391 -----GAAGCGTCGATATAT 9407
DB 2344 nGluAlaValTyrGlyAspAspLleSerAspSerPheValAspLysLeuGlnGluLe-- 2363
QY 9408 AAGCAACCTGATGTCAGATGATAT-----AGAAATATAT 9446
DB 2364 -----LysAsnAspValAspLleLleLysThrGlnAlaMetLysTyrG 2378
QY 9447 TACCTTGAAATATTTGCTAAAAAATACAAATGAAATCTAAATATATTAAGCAAT 9506
DB 2378 yMetLleCysLleProGluSerAsnPe-----ValGlyAlaSerTyrLleAspAl 2395
QY 9507 AAAAGATCAATCTCA-----GCTAATATAGAC----- 9534
DB 2395 aGlnAspMetSerAspLleValValLeuSerAlaSerTyrTyHisAsnLeuMetSerPr 2415
QY 9535 -----ATAAACCATCTGAAGAAATGTTCACTATATTAATCAAA 9578
DB 2415 oGluAspMetLysTyrAsnGlyValSerAsnGluGlnGlnAlaGluValLysLysAr 2435
QY 9579 AGATTCGAATGCGCTTGGAGTTAAATGAT-----AT 9611

```

Db 2435 GGLYLeuGInlLeaLeuThrThrLysGluAspLysGlyGlnAlaThrAlaSerLys 2455
OY 9612 AAATGAATAGTATGAGGACAAATAATGAATGAATGAATGATC----- 9657
Db 2455 SHISGLuTyValSerHisLysLeuAsnHisLysSerThrValSerThrLysSerG1 2475
OY 9658 -----AAGAAGTACTATAAAAAATATATCCGTGTTATATTTGTTGAGATGAACACA 9713
Db 2475 YAlaLysGlyLeuAlaGlnAlaAlaAlaThrThrAlaTyGluAspSerGlnSerH1 2495
OY 9714 CAAAAATCAGTACTAGATGCAATATATAAAGAAGACCAACAGCTGCTTAAGC 9773
Db 2495 SPRO-----GlnIleGluGlnSerHisArg----- 2504
OY 9774 ACTCTATTCTTACACCCCATGATGATTTCTTATCAACAGCCTTATTTCAACACA 9833
Db 2505 -----ThrAsnHis-----HisLysHisH1 2511
OY 9834 TCGAGTAGCACATATGATCTCTATAAATGATATTTGAAAGTAGATCTCTGT----- 9888
Db 2511 SLysArgGlnGlnSerLeuAsnSerHisSerThrSerLysThrHisSerSerArgAs 2531
OY 9889 -GTTATGTTATCGCGCTTATGAGTTGATGAGCGCTTCAATTCATGAGAAGAA----- 9936
Db 2531 nThrProAlaSerArgArgAspIleValAla-----SerPheMetSerArgAlaGlySerAl 2550
OY 9937 -AAATTCAAATCGCTGTGAGACTGTTGGCTATGATGAAATATCCGCA----- 9984
Db 2550 ASerArgThrAlaSerLeuGlnThrLeuAlaSerLeuAsnGluProSerIleIleProAl 2570
OY 9985 -----GGAGATGATGAATGCTTACGTAGTGGATGGAATCCAAATATGCTA 10025
Db 2570 AleuThrGlnThrValIleGlyGlnLys-----Leu-PheLysTrpTrpProAlaGlnGlyLys 2589
OY 10026 CATACCATATAGAGTGTCCATATATAAGCAAAACATATATATATATG 10075
Db 2589 rOpheGlyPheGlnSerArgHisGluArg-----PhePheTrp 2601

RESULT 13
MST2_DROHY
ID MST2_DROHY STANDARD; PRT; 1391 AA.
AC 008696;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Axoneme-associated protein mscl01(2).
GN MScl01(2).
   Drosophila hydei (Fruit fly).
   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
   NCBI_TaxID=7224;

OC
RN
RX
RA MEDLINE=95045538; PubMed=7957199;
RT "Tandemly arranged repeats of a novel highly charged 16-amino-acid
   motif representing the major component of the sperm-tail-specific
   axoneme-associated protein family Dmscl01 form extended
   alpha-helical rods within the extremely elongated spermatozoa of
   Drosophila hydei."
RL Eur. J. Biochem. 225:1089-1095(1994).
CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCTYTES AND EARLY
   SPERMATIDS
CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC -1- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
   STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
   REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

DR EMBL: X73481; CAA51876.1; -
 DR PIR: S34154; S34154.
 DR Flybase: FBgn020733; Dhyd\mscl01(2).
 KW Sperm; Repeat; Multigene family; Polymorphism.
 FT 59 x 16 AA APPROXIMATE TANDEM REPEATS OF
 FT DOMAIN 332 1268 [KRI-K-X-C-X-X-A-K-X-X-K-X-X-E.
 SQ SEQUENCE 1391 AA; 159000 MW; 1B2A368F30E4878 CRC64;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	4.23e-08	319.50	344	230	631	554
Percent Similarity:	32.63%					
Best Local Similarity:	19.56%					
Query Match:	1.64%					
DB:	1					80

US-10-087-013-1 (1-10628) x MST2_DROHY (1-1391)

```

OY 4519 GAAAAATGTAAGTCAAAATGACCAATGTAAGTGTATACAGATGTAATGAATGCGAG 4578
Db 50 GlnSerThrAspAspSerAsnGlnIleArgCysAspAlaAsp-----LysThrAlaAla 67
OY 4579 GACTACGTTAAATATATATGTAAGAAAAAAGAGTGCATTCACAAAGATTAATATTACAG 4638
Db 68 ASP-----LysLysLysGlnLysGln-----Lys 75
OY 4639 GATGACCGCAGCAAAAAAAGATTCGATGACAAACATGTTGTAATGCTTACAGACTAT 4698
Db 76 GluGlu-----GluGlnAlaLysIle-----ArgLys 85
OY 4699 ACTGCAACGAATGC-----AACAGATTAATTGAAACAGAAATTTACGCTAGT 4746
Db 85 YLysArgGluCysLeuLysValGlnLysArgValIleAlaGlnIleIleArgCys-Ser 104
OY 4747 TGTGCTGATTAAGCTCGAAGTGCCTGTGTGTCACAAAGATATATACATTTAGAAAAA 4806
Db 105 -GlyGlnLys-----AspArgIleLeuIleGlnIleGlnLysMetLysCysLeuThr--- 120
OY 4807 CAGCTTACTATGATGCGCAGCAACATTTGCGGCACAAATTTATGAAATGAC--- 4863
Db 121 -----AspGlyMetLysLys---AlaCysThrIleAlaLysAlaLysLeu 135
OY 4864 -----GACAAA-----TATACTAACATTTTCGATTAAGTAAAGTGCAGAA 4902
Db 136 IleAlaAspLysGluLeuAlaValGlnCysAlaAlaAlaLeuSerLysLysAspLysValLys 135
OY 4903 GGATTACTA-----AAGAGCGCAACACAGAGTGC 4932
Db 156 AlaleuLeuLysCysGluArgLysSerLysGlnLysCysLysAsnGlnAsnSer 175
OY 4933 -----ATTAAGTGGCAAAACAAAGAGTCTTATTAATAC 4965
Db 176 ProAlaGluCylAspLysAspArgThrLysLysGlyLysThrLysGlyLysSerGlyGly 195
OY 4966 AATPACTTGAAGAATGACTGAAGAT-----GTGCTTTTCT 5004
Db 196 GlyAsnLysLysArgSerThrLysGluAsnArgAlaLysLysGlyLysLysLeuValLys 215
OY 5005 TCTGCTGACGACTATATGTTTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 5043
Db 216 AsnArgPheThrGlnLysLeuGlnLysCysIleLysSerGlnIleAlaAspValCysGln 235
OY 5044 -----AATTATACAGATCCAGAA----- 5061
Db 236 CysArgGlnAsnPheThrGlnLysAspGluArgLysArgGluAlaAlaSerTrpLysCysMet 255

```

QY	5062	-----GTTAAATCAATAAATGGCTGCAAAAAAGTTGATG-----GAAGTCGGCCAGCAACGAA	5115
Db	256	GLYthrlYsLIeYsSerLIeCYsgrLYsAlYvalIIleAlglImeCYsGLuAlaIa	275
QY	5116	GGGTACAAATTGGGTCAATACTACAAAGAAAAAGAAAAAGAAAAAGAAAAATAAATAAACCTGC	5179
Db	276	GLYthrlYsLIeYsSer-----SerGLuProlYsLYsGLYLYsLYsLYsLYsAsn	292
QY	5176	GATGCGCAAAATATCTTATGAGTGCCGCCCTGTAGTGTATGAATATATAGTTTTAT	5235
Db	293	AspGLuLYsLYsGLuLYs-----	298
QY	5236	GATTTAAGATATTAATTCAGTATATGATTAATTGGAGAGATGA-----AAACAAG	5289
Db	299	-----GluLeuGluArgGLuIIleuLeuLYsGLuIn	308
QY	5290	ACCAGCAAAATTTGAGAAAAATATTTAACAAAAAATGCAACATCAGTTGGCAACGAAGT	5349
Db	309	AlaGLuInuLIuAlaLYs-----	314
Db	315	-----IleArgGLYvalIValLYsGLuValLYsLYs	324
QY	5410	GAATGTGTGGAACCGAATGATATCCGGGTACAAAACGTGGTAGGATGATGAAATAGT	5466
Db	325	LYsCYsLYsGLuLYsAlaLeu-----LYsLYsCYsLYsAsp-----Leu	338
QY	5470	GGAATATGTGCAGAAAGTGTATGAGATCTTAAAAAAATGCGTTCGTACCTTCAGATGAT	5529
Db	339	GLYthrlYsMeLIeLYsGLuInaGLuLYsLYsCYsAlaIaLeuAlaLYs-----	356
QY	5530	GATTATTCATGGGAAAAATCCGCATGAAGTACTGCGTATCAGTTCTTCGATGATT	5589
Db	356	-----	356
QY	5590	GCCGATGGGGTGAAGATTTTTGCACAACTAAAGAAAAAGAAATTGGACAAATTGCTACGG	5649
Db	357	-----LYsGLuLYsGLuInaLYsGLuLYsLYs-----	365
QY	5650	GCGGTATGATTAATCTTGTGTGATTAATGA-----GATTAAGAAAGAAATCTACAGT	5706
Db	366	AlaCYsLYsGLuLeuAlaLYsLYsLYsGLuAlaAspLIuLYsLYsCYsGLuInu	385
QY	5707	GCGGTACACAAATATTAATTTATTAGTGAAGTGAACCAACAGTATGAAAAAATC	5766
Db	386	AlaAlaAsnLYsGLuLYsLYs-----AlaIaGLuLYsLYsCYsGLuLYsAlaIa	403
QY	5767	AAAAAATATGTGAGATTAAGCAAAATATATTCGAGCATCTGTGGCAAAAGATCA	5828
Db	404	LYsGLuInuLYsGLuAlaIaGLuLYsLYsCYsGLuGLuAlaIaLYsLYsGLuLYs	423
QY	5827	GAGGAGCCTGCG-----GAATATTATGACAAACAATTAATAAAAAATTGTGA	5874
Db	424	GLuAlaIaGLuArgLYsLYsCYsGLuInuLeuAlaLYsAsnLIeLYsLYsAlaIaGLu	443
QY	5875	AATTAAGTGAAGATTGTGAATATATAGTATGAGAAAGATGTGCACACAGCATTAAT	5934
Db	444	LYsLYsLYsCYsLYsGLuAlaIaLYsLYsGLuLYsGLuAlaIaGLuArgLYsLYsCYs	463
QY	5935	GATGTGTAATAGTCMAAATATATGCCCCGATCATTTAGACAGATGAACCAAAAGATGAAGA	5994
Db	464	GLuInuLeuAlaLYsLYsLIeLYsLYsAlaIaGLuLYsLYs-----	477
QY	5995	AAGGTATATTGTCAAGTCCAGGAGGTCACCAACGCTGTACGAAGGGAAACACCGTCACCA	6055
Db	478	LYsCYsGLuInuThrlAlaLYsLYsGLYLYsGLuValAlaGLuArgLYsLYsCYsGLuInu	497
QY	6055	CGGGTATCACTGATATCAAAAGCAGCGCATCG-----AAAAAGAACGCAAAACA	6105
Db	498	LeuAlaLYsLYsLIeLYsLYsAlaGLuIIleLYsLYsCYsLYsLYsLYsLeuAlaLYsLYs	517
QY	6106	GCGCCGCTTACAAAACGCCGAAAAAAGTGGAAATCTTAACAACAGAAATCGCAGCACAA	6165

[illegible]

QY 8215 -----GATCACCTGTGATGATTATATACCTCAAGGTTGGATGATGAT 826
 Db 1037 LysLysLeuLysGluAlaAlaGluLysLysGlnCysGlnGluAlaLysLysLeuLys 1056
 QY 8263 GAATGGCTGCAATATATATATGTTAAAGCACTGATGAGAAATTTGAAAAATCA 8322
 Db 1057 GluAlaAlaGluGln-----LysGlnCysGlnGluAlaLysLysLeuLysGluAla 1074
 QY 8323 TGTGATCACTGTTAAACAATCTGACAGATGCAAGAAATATATATATATATAGTGA 8382
 Db 1075 AlaGluLysLysGlnCysGlnGluAlaLysLysGlnLysGluAlaAlaGluLysLys 1096
 QY 8383 CAGCTGTTAAACGAGATGTCAGAAATATATATATTTCTTTAAATGAAATCTCATTC 8442
 Db 1095 GlnCysGlnGluAlaLysLysGlnLysLeuLysGlnLysAlaGluLysLysGlnCys 1112
 QY 8443 GATATATCAATCAAAATATATATACAAAGAAATGTATGACACACCAATATATCAAAATCTCT 8502
 Db 1113 GlnGluAlaAlaLysLysGlnLysGlnLysAlaAlaGluLysLysArgCysGlnLysAlaAla 1132
 QY 8503 ACTTATGATCATGTTCCAAATTTTGTATCAAAAGTTGCAAACTTTTAAAGTAAATGTTCT 8562
 Db 1133 LysArg-----GlnLysGluAlaAlaGluLysLysCysAla 1145
 QY 8563 GTTGAAGCTTTTCTGGA---TATCTTCATGAAACAGATAGTGTGTAATTAATTT 8619
 Db 1146 GluAlaAlaLysLysGlnLysGlnLysAlaAlaGlnGlnLysGlnLysCysAlaGluAlaLys 1165
 QY 8620 AATGAAATATGATGGTCTTCACATATATCAATATGCTTTGGAAGAACACCAAAAGT 8679
 Db 1166 LysGlnLysGlnAlaAlaGluLysLysLysCys-----AlaGluAlaAlaLysArg 1182
 QY 8680 TATAAGAACCTGC-----AGTTGACACTCTCTTAAGAAATCCATTTGCATAT 8730
 Db 1183 GlnLysGlnAlaAlaGlnLysLysCysAlaAspLeuAlaLysGlnGlnGln----- 1201
 QY 8731 TGTCTTACCGATCAAAACCAAGATGATGATGAGATTTACAACTTTTACTTCTGCTCG 8790
 Db 1202 ---ProAlaGlnMetLysLys-----CysGlnGln-----AlaAla 1212
 QY 8791 AAGAAATGATTTGATTAATTAATCTTGATTAATGAGACATCTGTTTAAATAGTTCA 8850
 Db 1213 LysLysGlnLysGln----- 1217
 QY 8851 GATGATTAACAAGGTGATTAATGATTCCTCCAGAAGAAAGACATTTATGTACAAGCTATC 8910
 Db 1218 -----AlaAlaGlnLysGlnLysCysAlaLysAla-- 1227
 QY 8911 ACTGCATTAATTAATAGAAAAGGTGATTAAGAATTTTAAAAAAAACCTTCACTACTCT 8970
 Db 1228 -----AlaLysLysGlnLysGlnAlaAlaGlnLysLysCysAlaGln 1242
 QY 8971 GCTTTCAGTCAAGCAATTTAGTTCAAAAAATTAATATGGAAGAAGCTTGCTTT 9030
 Db 1243 AlaAlaLysLysGlnGlnGlnGlnAlaAlaGlnLysLysCysAlaGln- 1258
 QY 9031 GAGCAATGAATAATACTGATTCAGATTTATTCAGATTAATTAAGAAGCAATGATGATG 9090
 Db 1259 AlaAlaLysLysGlnLysGlnAlaGlnLysLysArgLysCysGlnLysAlaGlnLysAla 1278
 QY 9091 GACACTCATTTCTGAAAAAATTAATAAATAATTTGAAACATCAATGAAGCAACGAA 9150
 Db 1279 ---AlaAlaLeuLysArgLysCysAlaLysLeuValIleArgAlaLysGlnAlaAlaLeu 1297
 QY 9151 AATGTATAAACAATGAGGGAATAATATAGCTGATATGCAACGCTATGTTATGGA 9210
 Db 1298 ArgLysLys-----CysAla 1302
 QY 9211 TATTAATTTCTACTTCAAAAGTAACTATGATGAGAGATGCTGTCAATTACCAAAAGAT 9270
 Db 1303 IleIleAlaLysLysAlaLysMetAlaAlaGlnLysLysGlnCys----- 1317

[illegible]

Db 1765 uLysAlaGlnAspLeuLysIleGlnGluLeuArgIleAlaHisMetHisLeuLysG1 1785
OY 6201 ACAGACATCAACAGCAACAACAACAGAAATGTGACGTGGGACAAATGGTAAAGGCCATTTCT 6260
Db 1785 uGlnGlnGluThrIle-----AspLysLeuArgIleVal 1797
OY 6261 TTGGATTAACACGATAGCAGGGGGGTGATAGAGGTTGTAATCCAAAACGATAGGACA 6320
Db 1797 IserGluLysThrAspLysLeuSerAsnMetGln----- 1808
OY 6321 ATATCCTAATGGGTTGTATGTAGTAAAGTAAGAAATGAAATGCAATATGATAT 6380
Db 1809 -----LysAspLeuGlnAsnSerAsnAla---LysIle 1818
OY 6381 GCCCTCAGAGAAAAAATTATGTATTAATATATACAAATATTAATTAATGAAATGCA 6440
Db 1818 uGlnGluLysIleGlnGluLeuLysAlaAsnGlnHisGlnLeuIleThrLeuLysLysAs 1838
OY 6441 A---AATAGCGTGACAAATGATTAAGAGCGTTTATTAATGTCAGCAATAGCAAC 6497
Db 1838 rValAsnGluThrGlnLysLysValSerGluMetGluGlnLeuLysGlnIleLysAs 1858
OY 6498 TCATATTTTGTGGTTA---AAATATATTAATGAAATCCTGCAGCAAAATGCAATGCA 6554
Db 1858 pGlnSerLeuThrLeuSerLysLeuGluIleGlnAsnLeuAsnLeuAlaGlnGluLeuH1 1878
OY 6555 AATGGAACAATTCAGATGATTAATAAGAAATATGTATTAATGTCATATGATTAATA 6614
Db 1878 sGlnAsnLeu-----GlnGluMetLysSerValMet----- 1888
OY 6615 AGATATGTTTTTGGAACTGATATTTCTAATGATTAATAAAATTAATACGTAAACAATAG 6674
Db 1888 ----- 1888
OY 6675 TGTAAACACCATTCATGATAAATAATAGAAAAACAGATAAAAAAGATAGAGA 6734
Db 1889 -----LysGlnArgAspAsnLeuArgValGlnG1 1899
OY 6735 ATTACGTAATATTTTGGAGAAAAAATTAATTTATTTGGGAAGCAATATATATG 6794
Db 1899 uThrLeuLysLeuGlnArgAspGlnLeuLysGlnSerLeuGlnGluThrLysAlaArgAs 1919
OY 6795 ATTA-----ACTTATCATCTCACAGACAGAAAAAGAAAGAA 6831
Db 1919 pLeuGlnIleGlnGlnGlnLeuLysThrAlaArgMetLeuSerLysGlnHisLysGlnH1 1939
OY 6832 -----AAATTAAGAT-----AATTAACAGTACATGACATGACCA 6869
Db 1939 rValAspLysLeuArgGlnLysIleSerGlnLysThrIleGlnIleSerAspIleGlnL1 1959
OY 6870 A-----CTGACGCGCTCCCTTGAAAGGTTTGTAAAAAGGCC 6905
Db 1959 sAspLeuAspLysSerLysAspGlnLeuGlnLysLysIleGlnGlnLeuGlnLysLysG1 1979
OY 6906 C---CAATTTTGAAGATGTTTCACAGAAATGGGACAGAAATTTGTATAAGAGAGAA 6962
Db 1979 uLeuGlnLeuLeuArg-----ValLysGlnAspValAsnMetSerHisLys 1994
OY 6963 ACAAGTTGTAATTAATGAGGGCGGCTGAAGCATATAG-----TG 7004
Db 1994 sLysIleAsnGlnMetGlnGlnLeuLysLysGlnIlePheGlnProAsnTrpLeuCysLysC1 2014
OY 7005 TAAATGTTAGTATGACGCTAGACACACAGAAATGTCAGAGCGCTGTAAACATATCAAA 7064
Db 2014 sGlnMetAspAsnIleGlnLeuThrLysLysLeuHisGlnSerLeuGlnIleArg-- 2033
OY 7065 TTTTATTTAAGAGTGGAAAACTGAATATGAAAGCAAAAGAAAGTCAAAAAGATAA 7124
Db 2034 -IleValAlaLysGlnArgAspGlnLeuArgIleLysGlnSerLeuLysMetGlnArg 2053
OY 7125 AGATGGCAAAAGTAAAGATATCTCTTACTAGAAAGACATA----- 7170
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

Db 2053 gAsp-----GlnPheIleAlaThrLeuArgGlnMetIleAlaArgAspArg 2068
OY 7171 -----GAGAGGCAACATGCTCT-----CATGAATATTAA 7202
Db 2068 gGlnAsnHisGlnValLysProGlnLysArgLeuLeuSerAspGlnGlnGlnHisLeuH1 2088
OY 7203 CATGAATTAAGAAATTTGTGGCAATTAAGATGTTCTGTATAGCAAAACCTTTCT 7262
Db 2088 tGlnSerLeuArgGlnLysCys-----Se 2096
OY 7263 ACAACATACAAAAACAACAACAACAATCACAATCATCAGTATGCTAATGATAGCAGAAATC 7322
Db 2096 rArgIleLysGlnLeuLeuLys-----ArgTrpSerGlnMetAspAsnHisLysGlnC1 2114
OY 7323 GCTGATATGTTCTCGAAGATTTTAACAAG---TGTAAGTGTCCGTGAACCTTTCAAAAA 7379
Db 2114 sLeuAsnArgLeuSerLeuAspLeuGlnLysGlnIleGlnPheHisArgIleMetLysL1 2134
OY 7380 GGGATCTATGAT-----CATACAAAAAATTAAGTAACTAAATACCTATGATTTG 7433
Db 2134 sLeuLysTrpValLeuSerTrpValThrLysIleLysGlnGlnGln-----HisGlnC1 2152
OY 7434 TGTAGAGAAAGCAGCATATTTATTTATTAAGAACACAGAAATTAATGATATTTACCT 7493
Db 2152 sIleAsnLysPheGlnMetAspPheIleAspGlnValGlnLysGlnLysGlnLeuH1 2172
OY 7494 GAAGAAAAATTTATACCTATGAG---TGTACAAAGAAAAAGAAATTAATTAATGATTTG 7550
Db 2172 eLysIleGlnHisLeuGlnGlnGlnAspCysAspValProSerArgGlnLeuArgAspLeuL1 2192
OY 7551 GACTAATAATTAACCTTGC-----GATCCTAAGAACCTTTGACACCTGATTA 7598
Db 2192 sLeuAsnGlnAsnMetAspLeuHisIleGlnGlnIleLysAspPheSerGlnSerG1 2212
OY 7599 ATATATAGAAAGAAACCTTGTGAATATAGAAATAAGAAATAAGTTTAAGATAGATTA 7658
Db 2212 uPheProSerIleLysTrpGlnPheGlnGlnValLeuSerAsnArgLysGlnMetThrG1 2232
OY 7659 TGAATGAATGTTTCAAAAATTCAAAGTCTATCAGGAGAAAAAGATAGATTTGATAC 7718
Db 2232 pPheLeuGlnGlnTrpLeuAsnThrArgPheAspIleGlnLysLeuLysAsnGlnIleG1 2252
OY 7719 TCACAAAGAGACAAACATAGTCTTAAGAAAT----- 7749
Db 2252 nLysGlnAsnAspArgIleCysGlnValAlaAsnAsnPheAsnAsnArgIleIleAla11 2272
OY 7750 -----TTAGATGAATTAATTAAGTAAAGCTTAAAGTAAATTA 7790
Db 2272 eMetAsnGlnSerThrGlnPheGlnGlnArgSerAlaThrIleSerLysGlnTrpGlnG1 2292
OY 7791 TCTCCTAAAAATGGTGTGCGAAGTGCACAGAAATGAAGAAATAGACATTAATAAAACCTT 7850
Db 2292 nAspLeuLysSerLeuLysGln-----LysAsnGln-----LysLeuPheLysAsnTr1 2308
OY 7851 CAAGCTCA-----GAGAACGGGTGGCGCAATGATCAATATGTGATTAAT 7895
Db 2308 rGlnThrLeuLysThrSerLeuAlaSerGlnArgIleValAlaAsnAsnTrpThrGlnAspAs 2328
OY 7896 GAAATATAGTTTCCGTGATGCGGTGACATATAGAGAGAAAGATGTTTAAAGAAATGG 7955
Db 2328 nLysAsnProHisValThrSerArgAlaThrGlnLeuThrThrGlnLysIleArg----- 2346
OY 7956 TGGTACTTACCTCCCTGACAAATTAATTAATATAGGTTTGAATATATATAGGAAA 8015
Db 2347 -----GlnLeuGlnAsnSerLeuHisGlnAla----- 2355
OY 8016 ATGAGAAATAAATTAAGGTAGAAATTAATTAACAGATGACAAACGTTGCTGTGC 8075
Db 2356 -----LysGlnSerAlaMetHisLysGlnSerLysIleIleLysMetGlnLysG1 2372
OY 8076 TTGGTGGGATGCTAATTAAGAAATATTTGAAAGCAATATACGTGCAACACAGACAGA 8135
Db 2372 uLeuGlnValThrAsn-----AspIleIleAlaLysLeuGlnAlaLysValHisGlnLe 2390
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

QY 8136 TCGAAACCTTTTTCAGAAAGGAAGATGCGATGATTGAACGCAATTAATACACAGA 8195
 DB 2390 rAsnLysCysLeuGluLysThrLys-----GluThrIleGlnValLeuGlnAs 2406
 QY 8196 TAAGTTGGACATACAGACAGATCCACTGTGATGATTAATATCCCAACGGTTTGATG 8255
 DB 2406 pLysValAlaLeuGlyAlaLysProTyrLysGluGluIleGluAspLeuLysMetLys 2426
 QY 8256 GAGG-----ACTGAATGCTCTGAATATTATTGTAAGACAGCTAGTGAAGAATTGA 8259
 DB 2426 uValLysIleAspLeuGluLysMetLysAsnAlaLysGluPheGluLysGluIleSerAl 2446
 QY 8260 -----ACTGAATGCTCTGAATATTATTGTAAGACAGCTAGTGAAGAATTGA 8306
 DB 2446 arThrLysAlaThrValGluLysGlnLysGluValIleArgLeuLeuArgLysLysLeuAr 2466
 QY 8307 AAAATTTAAAAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8366
 DB 2466 gATGSerGlnGlnAlaGlnAspThrSerValIleSerGlnHis--ThrAspProGlnPr 2485
 QY 8367 TGAATAATAG-----TGTGACACAGTGTAAACGAGATGTCAGAAATATATATATTT 8417
 DB 2485 oSerAsnLysProLeuThrCysGlyLysGlySerGlyIleValGlnAsnThrLysAlaLe 2505
 QY 8418 TGTCTTTAATGGAATCTCTA---TTCGATATACATCAATTAATACAAAGAAATTGTA 8474
 DB 2505 uIleLeuLysSerGlnHisIleArgLeuGluLysGluIleSerLysLeuLysGlnGlnAs 2525
 QY 8475 TGAAACAACCAATATATACAAA-----ATCTCTACTTATGATGATGATGATGATGATG 8525
 DB 2525 nGluGlnIleLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnL 2545
 QY 8526 TGTACAA-----AGTGTGAAACCTTTTAAAGTGAATGCTCTGTGAGAGCTTTTCTGA 8579
 DB 2545 uValLysThrIlePheLysGluArgThrLeuLysArgGluAla----- 2588
 QY 8580 ATATCTTCATGAAACAAGTAAGTGAATTAAT---AAATTAATGAAATGATGCTTC 8636
 DB 2559 -----HisLysGlnValThrCysGlnLysSerProLysSerProLysValThrGlyTh 2576
 QY 8637 TTTCT-----AATTAAGCAACATA 8654
 DB 2576 rAlaSerLysLysLysGlnIleThrProSerGlnCysLysGluArgGlnLysLeuIleAsp 2596
 QY 8655 TCGTTTCGAGAAACACCAAAAAGT---TATAAAGAGCTTGCAGTGTGACTACTCTTC 8711
 DB 2596 oValProLysGlnSerProLysSerCysPhePheAspSerArgSerLysSerLeuProse 2616
 QY 8712 TAAGAAATCCATTG-----GATTAAT-----TGTCTTACGATCAAAA 8747
 DB 2616 rProHisProValArgTyrPheAspAsnSerSerLeuGluLysCysProGluValGlnAs 2636
 QY 8748 C 8748
 DB 2636 n 2636
 RESULT 15
 ATRX_HUMAN
 AC P46100; P51068; Q15886; Q9NRS3; Q9H0Z1;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked
 DE nuclear protein) (XNP) (Znf-HX).
 GN ATRX OR RAD54L OR XH2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANT S-1860, AND

RP VARIANTS ATR-X.
 RX MEDLINE-97123494; PubMed-8968741;
 RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
 RA Gibbons R.J.;
 RT "ATRX encodes a novel member of the SMF2 family of proteins: mutations
 RT point to a common mechanism underlying the ATR-X syndrome.";
 RL Hum. Mol. Genet. 5:1899-1907(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
 RX MEDLINE-97386582; PubMed-9244431;
 RA Villard L., Iossi A.-M., Cardoso C., Proud V., Chiaroni P.,
 RA Colleaux L., Schartz C., Fontes M.;
 RT "Determination of the genomic structure of the XNP/ATRX gene encoding
 RT a potential zinc finger helicase.";
 RL Genomics 43:149-155(1997).
 RN [3]
 RP SEQUENCE OF 860-2492 FROM N.A.
 RX MEDLINE-95179111; PubMed-7874112;
 RA Stayton C.L., Dabovic B., Gullisano M., Geez J., Broccoli V.,
 RA Giovanazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
 RA Bianchi M.E., Consalez G.G.;
 RT "Cloning and characterization of a new human Xq13 gene, encoding a
 RT putative helicase.";
 RL Hum. Mol. Genet. 3:1957-1964(1994).
 RN [4]
 RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE-94214473; PubMed-8162050;
 RA Geez J., Pollard H., Consalez G., Villard L., Stayton C.L.,
 RA Millaesau P., Khrestchatsky M., Fontes M.;
 RT "Cloning and expression of the murine homologue of a putative human
 RT X-linked nuclear protein gene closely linked to PKGI in Xq13.3.";
 RL Hum. Mol. Genet. 3:39-44(1994).
 RN [5]
 RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
 RX MEDLINE-95211835; PubMed-7697714;
 RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
 RT "Mutations in a putative global transcriptional regulator cause X-
 RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
 RL Cell 80:837-845(1995).
 RN [6]
 RP SEQUENCE OF 1375-2492 FROM N.A.
 RA Pearce A., Chapman J.;
 RL Submitted (Dec-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP EZH2 BINDING.
 RX MEDLINE-98167853; PubMed-9499421;
 RA Cardoso C., Timsit S., Villard L., Khrestchatsky M., Fontes M.,
 RA Colleaux L.;
 RT "Specific interaction between the XNP/ATR-X gene product and the SET
 RT domain of the human EZH2 protein.";
 RL Hum. Mol. Genet. 7:679-684(1998).
 RN [8]
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
 RP HETEROCHROMATIN.
 RX MEDLINE-20040663; PubMed-10570185;
 RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
 RA Bickmore W.A., Pombor A., Turley H., Gatter K., Picketts D.J.,
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
 RT "Localization of a putative transcriptional regulator (ATRX) at
 RT pericentromeric heterochromatin and the short arms of acrocentric
 RT chromosomes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
 RN [9]
 RP DISEASE.
 RX MEDLINE-20213147; PubMed-10751095;
 RA Villard L., Fontes M., Ages L.C., Geez J.;
 RT "Identification of a mutation in the XNP/ATR-X gene in a family
 RT reported as Smith-Fineman-Myers syndrome.";
 RL Am. J. Med. Genet. 91:83-85(2000).
 RN [10]
 RP VARIANT ATR-X SER-1713.
 RX MEDLINE-97196774; PubMed-9043863;
 RA Villard L., Lacombe D., Fontes M.;

RX MEDLINE=20123062; PubMed=10660327;
 RA Eichner M., Romano C., Castiglia L., Falla P., Ruberto C., Amata S.,
 RA Greco D., Cardoso C., Fontes M., Ragusa A.;
 RT "New mutations in XNP/ATR-X gene: a further contribution to
 RT genotype/phenotype relationship in ATR-X syndrome.";
 RL Hum. Mutat. 12:214-214(1998).
 RN [14]
 RP VARIANT SHS LYS-1742.
 RX MEDLINE=99347960; PubMed=10417298;
 RA Lossi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,
 RA Prieto F., Fontes M., Martinez F.;
 RT "Mutation of the XNP/ATR-X gene in a family with severe mental
 RT retardation, spastic paraplegia and skewed pattern of X inactivation:
 RT demonstration that the mutation is involved in the inactivation
 RT bias.";
 RL Am. J. Hum. Genet. 65:558-562(1999).
 RN [15]

VARIANT RMS THR-2050.
 MEDLINE-99326061; PubMed-10398237;
 Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,
 Curtis M.,
 "Carpenter Waziri syndrome results from a mutation in XNP.",
 Am. J. Med. Genet. 85:245-251(1999).
 [16]
 VARIANTS ATR-X E-175; 178-V--K-198 DEL, S-190; P-219; L-246 AND C-249
 MEDLINE-99219535; PubMed-10204841;
 Villard L., Bonino M.-C., Abidi F., Reguena A., Beloungue J.,
 Lossel A.-M., Sauter L., Bonnefont J.-P., Romano C., Fiehera M.,
 Lacombe D., Hanaauer A., Philip N., Schwartz C.E., Fontes M.,
 "Evaluation of a mutation screening strategy for sporadic cases of
 ATR-X syndrome.",
 J. Med. Genet. 36:183-186(1999).
 [17]
 VARIANTS ATR-X S-179; L-190; I-194; C-246; F-1552; S-1645 AND C-1847.
 MEDLINE-20451413; PubMed-10955512;
 Wada T., Kubota T., Fukushima Y., Satoh S.,
 "Molecular genetic study of Japanese patients with x-linked
 alpha-thalassemia/mental retardation syndrome (ATR-X).",
 Am. J. Med. Genet. 94:242-248(2000).
 -1- FUNKTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR, MODIFIES
 GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN
 BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.
 -1- SUBUNIT: PROBABLY BINDS EH2. BINDS ANNEKIN V IN A CALCIUM AND
 PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSELINE-DEPENDENT MANNER (BY
 similarity).
 -1- SUBCELLULAR LOCATION: NUCLEAR, ASSOCIATED WITH PERICENTROMERIC
 HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
 INTERACTING WITH HPL.
 -1- ALTERNATIVE PRODUCTS: 5 ISOFORMS, 1, 2, 3, 4 (SHOWN HERE) AND 5;
 ARE PRODUCED BY ALTERNATIVE SPLICING.
 -1- TISSUE SPECIFICITY: UBIOUITOUS.

Pred. No.:	1.05e-07	Length:	2492
Score:	311.00	Matches:	522
Percent Similarity:	29.87%	Conservative:	344
Best Local Similarity:	18.01%	Mismatches:	914
Query Match:	1.60%	Indels:	1120
DB:	1	Gaps:	130

US-10-087-013-1 (1-10628) x ATRX_HUMAN (1-2492)

QY	2062	GAACATATGAAATATTAATAACA-----	AAAAAAAAAGATTTTC	2103
Db	55	AspMetCetLusnsSerLysgLuGlugLylhrSerSerSerGlyLysSerSerSer		74
QY	2104	AATTTGGAAATATATAG-----	GACTATTGTGAGATGCATTA	2142
Db	75	GlySerSerArSerLysrArLysProSerLeValThrLysTyValGlnSerAsp		94
QY	2143	GAACCTTGTGTAGTACATTAAAGAAAGAACTGCCACGATATGTAAAGACATTAATCAAC	2202	
Db	95	GLuTyPProLeuAsp-----	AspGLuThrValAsn	104
QY	2203	GAA-----GCATGTGAAACATCCCATATGCACAACAAACCGCTGTGTAAACCTCGT	2256	
Db	105	GLuAspAlaSerAsnGLuLusnsSerGLuLusnAspIleThrMetGLnSerLeu---ProLys	123	
QY	2257	GGAGGC-----ACGCAACCCCACTATAATAATTAAGAAGAAATAGCACATCTTTAAAGG	2310	
Db	124	GLyThrValIleValIleValGlnProGluProValLeuAsnGLuAspLysAspAspGlySer	143	
QY	2311	ACTGCATACGAGGAAGACACGAAATCGTGTCTTCATTAATTTGAAGGAAGACACAGGA	2370	

Db 144 ProGluPheArgSerIysMetLysThrGluAsnLeuLysLysArgGlyGluAsp 163
QY 2371 GGATATATATAAGCGGGGTAGAGAAAGACTTCAGACAAATTATGTAGAAATATATG 2430
Db 164 GlyLeuHis----- 166
QY 2431 ATAAACATCTTAATGTATCTTGGTTTTCAAATGACCAATGTATGCGCAAGGCACA 2490
Db 166 ----- 166
QY 2491 GGTGATGTTATACAAACAGATTGTGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 2550
Db 167 -----GlyLeuValSerCysThrAlaCysGlyGlnGln-----ValAsnHis 180
QY 2551 ATGCGTAAGAT-----CACGAAGATGTTATATGCTCCTAGACAGACAT 2598
Db 181 PheGluLysAspSerIleThrArgHis-----ProSerLeuGlnValLeu 195
QY 2599 ATATGTACATCCCAATTGTGACATTTCAACACGATGATCACCACCTTAT----- 2649
Db 196 IleCysLysAsnLysPheLysThrLysMetSerAspAspIleSerArgAspSerAspGly 215
QY 2650 -----GTAATATGTT-----GATGATTTAGTT 2673
Db 216 MetAspGluGlnCysArgTrpCysAlaGluGlyGlyLysAsnLeuIleCysCysAspPheCys 235
QY 2674 AATAATCTCTTTTGGGGGAGTCTCTATACGAAATATGTAAGCAACAAAGATATA 2733
Db 236 HisAsnAlaPheCysLysLysCysLysLeu-----ArgAsnLeuGlyArgLysGluLeu 253
QY 2734 CGAATGTATTAAGAAAGATATAC-----CTAAAGGCCCAAGAAAGTA 2778
Db 254 SerThrIleMetAspIleAsnAsnGlnTrpTrpCysTrpIleCysHisProGluProLeu 273
QY 2779 ACTGACCCAAACACGACAGACATATCTGCGAGCTATACGATTATGCGAGATATA 2838
Db 274 LeuAsp-----LeuValThrAlaCysAsnSerVal-----PheGluAsnLeu 287
QY 2839 GGTGATATATTTGAGAGAGATCTCTGGGAAAGAAACGATGATGTAAGCTGCA 2898
Db 288 -----GluGlnLeuLeuGlnGlnHisLysLysLysLysValAsp 301
QY 2899 GGCATTTTGAA-----ACTGTTTGTGTATATATACATATGCA 2937
Db 302 SerGluLysSerAsnLysValTrpGluHisThrSerArgPheSerProLysLysThrSer 321
QY 2938 CTCGAAAGCAAGAAATGATAATATATATGATGATGCC----- 2976
Db 322 SerAsnCysAsnGlyGluGlnLysLysLeuAspAspSerCysSerGlySerValThrTyr 341
QY 2977 -----CCCAATATTTA-----AAATTGAGGGA 3000
Db 342 SerTyrSerAlaLeuIleValProLysGluMetIleLysLysAlaLysLysIleGlu 361
QY 3001 AATTGGGGAACCTATAGACCAAGTATGGAAGCCATGAATGTATATAAATAT 3060
Db 362 ThrThrAlaAsnMetLysSerSerTyrVal-----LysPhe 373
QY 3061 TTGAGAGTAAATGGGACACCAATCAACAAAGATGATTATGCGATATAGTATCAT 3120
Db 374 LeuLysGluAlaThrAspAsnSerGluIleSerSerAla----- 386
QY 3121 ACACCATTTGATGATTTATCCCAAAAATTAGATGATGACCGAAGTGGCAAGATGG 3180
Db 387 -----ThrLysLeuArgGlnLeuLysAlaPheLysSerVal 398
QY 3181 TACTGCAAGTGCAGAAAAGAGATATGATAAGTTGAGAGAGAGATGAGGTAG 3240
Db 399 LeuAlaAspIleLysLysAlaHisLeuAlaLeuGluGlnLysLeuAsnSerGluPheArg 418
QY 3241 GATAGAGTAAATGCTAGCAAGGCTGTACGAAAGAGAGATGCTGTACGAGTGCACA 3300
Db 419 AlaMetAsp-----AlaValAsnLysGluLysAspThr----- 429

QY 3301 GAAGCTTGTATGAAATATATATGATATATAGATTTATGAAAGAAACATGAAATATATA 3360
Db 430 -----LysGluHisLysValIle 435
QY 3361 TCAGATTAATACAAAGA-----TTACATGAACAAGCA 3393
Db 436 AspAlaLysPheGluThrLysAlaArgLysGlyLysProCysAlaLeuGluLysLys 455
QY 3394 CAATGCTGTGTATGATATAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 3453
Db 456 AspIleSerLysSerGluAlaLysLysLeuSerArgLysGlnValAspSerGluHisMetHis 475
QY 3454 AGAATGTATTAATTAATTTTGTGCGAATTAATTAATTAATTAATTAATTAATTAATTAAT 3513
Db 476 GlnAsnValProThrGluGlnGlnArgThrAsnLysSerThrGlyGluHisLysLys 495
QY 3514 AGTGCTACTGATGATGAATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 3558
Db 496 SerAspArgLysGluGlnProGlnTrpGluProAlaAsnThrSerGluAspLeuAspMet 515
QY 3559 -----TATGAATGTTGAGCATAT 3579
Db 516 AspIleValSerValProSerSerValProGluAspIlePheGluAsnLeuGluThrAla 535
QY 3580 CTCGATGATACAGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3639
Db 536 Met-----GluValGlnSerSerValAspHisGlnLysArgGly 548
QY 3640 AGTATGCTAGAGTACGAAATTAATATGCTTTAGATATA-----CCACAG 3667
Db 549 SerSerLysThrGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 568
QY 3688 GACCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3744
Db 569 AspAsnArgGlyGlyIleLysSerLysThrThrAlaLysValThrLysGluLeuTyrVal 588
QY 3745 AAA-----ACGAAAAAAAGCGAGAAAGAGATGATGATGATGATGATGATGATGATGATGAT 3792
Db 589 LysLeuThrProValSerLeuProAsnSerProIleLysGlyAlaAspCysGln----- 606
QY 3793 AATGATATCTTAAAGAAAACGATGAAAGAAACAAATAGATGATGATGATGATGATGATGATGAT 3852
Db 607 ---GluValProGlnAspLysAspGlyTrpLysSerCys---GlyLeuAsnProLysLeu 624
QY 3853 AATAGTATGATATCCGATTTGCAATGCGCAATTAATTAATTAATTAATTAATTAATTAATTAAT 3906
Db 625 GluLysCysGlyLeuGlnGlnLysAsnSerAspAsnGlnHisLeuValGlnAsnGluVal 644
QY 3907 -----CCTGCTGTGTATGCCCCCTAGA 3930
Db 645 SerLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3933
QY 3931 AGA----- 3933
Db 665 ArgArgProThrGlnThrAsnProValThrSerAsnSerAspGluLysAsnGluThr 684
QY 3934 -----CAAAATTATGCTGATCTTTGCGCAATGAT----- 3966
Db 685 ValLysGluLysGlnLysLeuSerValProValArgLysLysAspLysArgAsnSerSer 704
QY 3967 -----AATGAATTAATAAATTAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 3999
Db 705 AspSerAlaIleAspAsnProLysProAsnLysLeuProLysSerLysGlnSerGluThr 724
QY 4000 TTAAGAA-----GCTTCATCAAAATGCGACGACGAGA 4035
Db 725 ValAspGlnAsnSerAspSerAspLeuMetLeuAlaIleLysGlyValSerArgMet 744
QY 4036 ACATTTCTTCATGCTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4089
Db 745 SerHisSer-----SerSerAspThrAspIleAsnGluIleHisThr 759


```
OY 4090 -----GATTAAGATTAAAGAGCAAAATTCCTCCCGCATTTTGG 4131
Db 760 AsnHsLysThrLeuTyAspLeuLysThrGlnAlaGlyLysAspAspLysGlyLysArg 779
OY 4132 AGATCCATGTTCTACACATTTGGAGATATATAGATTTTATTGGAACGATATATCA 4191
Db 780 LysArgLysSerSerThrSerGlySer-----AspPhe-----AspThrLys 793
OY 4192 AAAGTCACTGGTAGGGA-----AGTAACTAAAGAGCAAAATGATTTCTCT 4239
Db 794 LysGlyLysSerAlaLysSerSerLelSerLysLysArgGlnThrGlnSerGlu 813
OY 4240 TTCAAAAATGGTGAACCAAAATCTCTTAATGAAAAACCCCAAGATGGTGACAGAA 4299
Db 814 SerSerAsnTyAspSerGlu-----820
OY 4300 CATAGTCATGAGATATGAGAGCTATGCTATGCTACTAGTAAATTTGGGCAAAAAA 4359
Db 821 -----LeuGlnLysGlnLelLysSerMetSerLysLelGlyAlaAlaArg 835
OY 4360 -----GATGATTT-----4368
Db 836 ThrThrLysLysArgLelProAsnThrLysAspPheAspSerSerGluAspLysHis 855
OY 4369 -----ACCGAAACTACGCTTACACACAGTCAAAATTTAGT-----4404
Db 856 SerLysLysGlyMetAspAsnGlnGlyHisLysAsnLeuLysThrSerGlnLysLys 875
OY 4404 -----4404
Db 876 SerAspAspAlaGluArgLysGlnLysArgLutThrPheSerSerAlaGlnLysThrVal 895
OY 4405 GACAAAAGCACCACTTTGGAGAAATTTGCCAANCA-----CCCAGTTTATACA 4455
Db 896 AspLysAspPheThrThrLelMetGlnLelAspArgLeuProLysLysGlnAlaSer 915
OY 4456 TGCTTACCGAATGTGACGACGACTATGCTATACACACAAAAATAT-----4503
Db 916 AlaSerThrAspLysValAspLysLeuSerGlyLysGlnGlnSerPheThrSerLeuGlu 935
OY 4504 TTGAGAGATGTGCAGAAAAATGTAATGCAATGACCAATTTGAAGTGTATACAGAAATG 4563
Db 936 ValArgLysValAlaGlnThrLysGlnLysSerLysHisLeuLys-----ThrLysThr 953
OY 4564 AATTAAGAAATGGAGAGCTACGTT-----AATATATGAAAAAATAAAAAA 4608
Db 954 CysLysLysValGlnAspLysLeuSerAspLelAlaGlnLysPheLeuLysLysAspGln 973
OY 4609 GAGTGTATCCACAAATTAATATTTACAAAGATGAACGCGCAAAAAAAGATTCGATAGA 4668
Db 974 -----SerAspLutThrSerGluAsp-----AspLysLysGln-----984
OY 4669 CAACACATTTGTGTATGTTAGTACACTAATCTGAAAGATGCAACAGATTTCTGAAAC 4728
Db 984 -----984
OY 4729 AGGAAATTTACTGCTAGTTGGTGATTAAGCTTGAGAGTGCCTCTGTGTACAAAGAAAT 4788
Db 985 -----SerLysLysGlyThrGln-----990
OY 4789 ATCAATTTAGAAAAACAGGCTTACTATGATGCCGCAACAAATTTGGGTGCACAAA 4848
Db 991 -----GlnLysLysLysProSerAspPheLysLys-----Lys 1001
OY 4849 TTTTATGAAAAATGACGACAAATATACTAACATTTTCAGTAAGATAAGTGCAAAAGATTA 4908
Db 1002 ValLelLysMetGlnGlnLysGlnLysSerSer-----1013
OY 4909 GTAAGAGGGAACACAGGCTGCTATTATAGTGCAAAAACAAAGTCTTAATATACAAAT 4968
Db 1014 -----AspGlyThrGln-----1017
OY 4969 AACTTGAAGAATGACTGAGATGTGCTTTTCTCTCTGCTGACACTACGATATATGTTT 5028
Db 1018 LysLeuProGluArgGlnGlnLysCysHisPheProLysGlyLysLysGlnLysAsn 1037
OY 5029 CATGCAATGGATGCAATATATACAGATCCAGAAATTAAGATGAATAATGGTTGCCAAA 5088
Db 1038 GlyThrThrAspLysGlnLysLysSerLysLysLysLysLysLysLysLysLysLys 1057
OY 5089 AGATTGATG-----GAAGTGGCGGCAACGGAAGGTTACAAATTTG-----5127
Db 1058 AspGlnLeuSerAspTyAlaGlnLysSerThrGlyLysGlyAspSerCysAspSerSer 1077
OY 5128 -----GGTCACTACTACAAAAAATAAAAAA-----5151
Db 1078 GluAspLysLysSerLysAsnGlyAlaTyGlyGlnLysLysArgCysLysLeuLeu 1097
OY 5152 -----GAAAAAGAAAAAATAAAACGTGCGAGCGCACAAATATCTTAT 5196
Db 1098 GlyLysSerSerArgLysArgGlnAspCysSerSerSerAspThrGlnLysLysMet 1117
OY 5197 GAGGTCCCGCTGTAGTCTATGAAATATAGTTTATGATTTAAGCATTAATCTTA 5256
Db 1118 LysGlnAspLysCysAsnSerSer-----AspLysArg-----LeuLys 1130
OY 5257 GGTATGTATATTTGCAAGATGAAAAACAAAGACCGAGAAATTTGAGAAAAATATTT 5316
Db 1131 ArgLelGlnLysArgGlnArgAsnLeuSerSerLysArgAsnThrLysGlnLys 1149
OY 5317 AACAAAAATGCAACATCGAGTGGCAAGAGAGTATGATCTACTACAGAAAT-----5367
Db 1150 -----GlnSerGlySerSerSerSerAspAlaGlnGlnSerSerGluAspAsnLysLysLys 1168
OY 5368 -----CCCGAGAGTACGCGCAAAATTTTCTGCAACGAAAAATTAAGAAATGTG 5418
Db 1169 LysGlnArgThrSerSerLysLysLysAlaValLelValLysGlnLysLysArg-----1186
OY 5419 TGGAAACGAATGATATCGCGGTACAAACGTGAGGATGTAAGTAAGTGAATAGT 5478
Db 1187 -----AsnSerLeuArgThrSerThrLysArgLysGlnAlaSerPheThrSerSerSer 1205
OY 5479 GCAAGA-----AGTGAAGATCTAAAAAATGTCGTCTGACTTCAATGATGAT-- 5532
Db 1206 SerAspLelGlnLysAspAspGlnAsnSerLelLysGlnLysSerSerAspGlnLys 1225
OY 5533 ---TATCTATGCGGAAAAATCGCATGAAGTACTGCTATGATGTTTCTTCGATGTTT 5589
Db 1226 LelLysProValThrGluAsn-----LeuValLeuSer 1236
OY 5590 GCCGAATGGGTGAAGATTTTTCGAAACATTA--GAAGAAGAAATGGAAGAAATTCGTA 5646
Db 1237 SerHisThrGly-----PheCysGlnSerSerGlyAspGlnAlaLeuSerLysSerVal 1254
OY 5647 GGGCGTGTATGATTTACTTGTGATATGAGATGAAGATGAAGAAAGAAATGTACAGAT 5706
Db 1255 ProValThrValAspAspAspAspAspAspAspProGlnAsnArgLysLysLys 1272
OY 5707 GCGGTGACAAATATTAATAATTTATGATGAGTGAAGAACACAGATGAAAAACAAATC 5766
Db 1273 -----LysLysMetLeuLys-----1278
OY 5767 AAAAAATATGTGAGATAAAGACAAATAATATTCGAGACATCCTGTGCCAAAGATGCA 5826
Db 1279 -----GlnLysLysAlaAsnLeuSerSerAspLysAspLysSerSerAsp---1293
OY 5827 GAGAGCGCTCGCAATATTTAGACAAACAAATTAATAAATTTTGTGAATAATAAGTGA 5886
Db 1294 ---AspGlnProGlnLysLysLysArgThrGlnLysGlnAsnGlnLysAsnProGly 1312
OY 5887 GATTGTGAATATATAGCTATGAAGATGTGTCACACAGCATTAACGTGATTAATAGT 5946
Db 1313 AspGlnGlnAlaLys-----AsnGln 1319
OY 5947 CAATAATGCCCGCATCATTAAGCATGAACCAAAAGAAAGTGAAGAAAGTGAATTTGT 6006
```

Db 1320 ValAsnSerGluSerAspSerGlu-----GluSerLys----- 1332
QY 6007 CAAGTCCAGCAGAGCTCCACCAGTGTACAGAGGAAACCGCTCACCAGCGGTATCACTG 6066
Db 1333 -----LysProArgTyrArgHisArgLeu-----Leu 1341
QY 6067 ATATCAAAAGCGCAGATCGAAA-----AAAGAGCGAAACAGCGCGCGCT 6114
Db 1342 ArgHisLysLeuThrValSerAspGlyGluSerGlyGluGluLysThrLysPro--- 1360
QY 6115 ACAAAACAGCCGAAAAATGTGAAAAATCAACACAAATCGCAGACAAACAGCAGCC 6174
Db 1361 ---LysGluHisLysGluValLysGly----- 1368
QY 6175 CGACGAGCAGCAGACAAACAGAAACAGACATCAACAGCAACAAACAGCAATGTGAC 6234
Db 1369 -----ArgAsnArgArgLysValSerSerGluAspSerGluAspSerAsp 1383
QY 6235 GTG---GGCACAATGTGTAAAGCCATCTTTCGAATTAACAGATAGCAGCGGTGGAATA 6291
Db 1384 PheGlnGluSerGlyValSerGluValSerGluSerGluAspGluGln----- 1400
QY 6292 GAGCGTTGTAAATCAAAAGCGATGACATATCTAAATGCGGTTGTATTTAGGTAAAG 6351
Db 1401 -----ArgProArgThrArgSerAlaLysLys-----AlaGlu 1411
QY 6352 TCTAAAGAAATGCAAAATGCGATATGTATGCTCTAGAGAAATAATTTATGTATTAAT 6411
Db 1412 LeuGlnGluAsnGlnArgSerTyrLysGlnLysLysArgArgLysValGln 1431
QY 6412 AATATCAATTTAATTAATGAAATGAAATTAACCGTGACATATTAAGAGCGCT 6471
Db 1432 Glu-----AspSerSerGluAsnLysSerAsnSerLysGluGluGluGlu 1447
QY 6472 TTTAATTAATGTGCACCATTAATCTATTTTGTGTTAAATATATTAATTAATTAAT 6531
Db 1448 GluLysGluGluGluGluGluGluGlu-----GluGluGlu 1460
QY 6532 CCTGCAGCAGAAATGCAATTCGAAATGCAACATTCAGAGTAATTAAGATTAATG 6591
Db 1461 GluGluGluGluGluGluGluGluAsnAspSerLysSerProGlyLysGlyLysLys 1480
QY 6592 TATTATACATATGCGATTAATAAGATATGTTTGAACGTATTTCTATATGATTA 6651
Db 1481 -----ArgLysLysLeuLysAsp 1487
QY 6652 AAAATTTAATCTGTACAAATAGTGTACAAACCATTCATGAATTAATAAGAAAAA 6711
Db 1488 LysLeuArgThrGluThrGlnAsnAlaLeu----- 1497
QY 6712 CAGGATTAATAAAAGATGAGATTAAGTAAATATTTGGAGAAAAATAAAATTTT 6771
Db 1498 -----LysGluGluGluGluGluGluGluArgGlyLysGluGluGluGlu 1510
QY 6772 ATTTGGGAAGATGATATATGATTAATCACTATCACTCAGACAGAAAAAGAAAGAA 6831
Db 1511 -----GluArgGluArgGlu 1515
QY 6832 AAAATTTAGATTAATTAACAGTACATGAC-----ATGACG 6867
Db 1516 LysLeuArgGluValIleGluIleGluAspAlaSerProThrLysCysProIleThr 1535
QY 6868 AAACTAGCGCTTCCCTTGAAGAGTTTGTAAAGAGCC----- 6906
Db 1536 LysLeuValLeuAspLysGluGluThrLysGluProLeuValGlnValHisArgAsn 1555
QY 6907 -----CAATTTTGGATGTGTTACA 6927
Db 1556 MetValIleLysLeuLysProHisGlnValAspGlyValGlnPheMet----- 1571
QY 6928 GAATGGCAGAGAAATTTGTATTAAGAGAGCAACAGTGTGTTAAATTTGAGCGCGC 6987
Db 1572 ---Trr---AspCysCysCysGluSerValLysLysThrLysLysSerProGlySerGly 1589

QY 6988 TGTAAAGATATGATGTATATGCTATATGACGGTAAAGACAGAAAGTGTGACAG--- 7044
Db 1590 CysIleLeuAlaHisCysMetGlyLeu-----GlyLysThrLeuGlnValLysPhe 1607
QY 7045 -----GCGTGTACATAT 7059
Db 1608 LeuHisThrValLeuLeuCysAspLysLeuAspPheSerThrAlaLeuValCysPro 1627
QY 7060 CAAATTTTATTAAGAATGCGAAAGCTGAATATTAAGAAAGCAAGAAAGAAAGTTCAAAAG 7119
Db 1628 LeuAsnThrAlaLeuAsnTrpMetAsnGluPheGlnLysTrpLysGluLysAsp 1647
QY 7120 GATTAAGATGGCAAAAGATTAAGATTAATCTTCTACTAGTAAAGACATAGAAAGCA 7179
Db 1648 AspGlu----- 1649
QY 7180 ACATGTCTCATGATATTTAAACATGAATTAAGAAATATGTGCAATTAAGATTC 7239
Db 1650 -----LysLeuGluValSerGluLeuAlaThrValLys----- 1660
QY 7240 TCTTGTATGCAAAACCTTCTTCACAATCTACAAAAACACAGCAATCAATCATCC 7299
Db 1660 ----- 1660
QY 7300 GATGCTAATGATATGCCAGAAATGCTGATTAATGTTCTCGAAGAAATTAACAAGTGTAG 7359
Db 1661 -----ArgProGlnGluArgSerTyrMetLeuGlnArgTrpGlnGlu----- 1674
QY 7360 TGTCTGAACTTTCAAAAAGGATCTATGATTCATACAAAAAATTAATGTAACCTTAA 7419
Db 1675 -----AspGlyLysValMetIle----- 1680
QY 7420 ATACTATGATTTGTGTAGAGAAAGCAGCATATTTATCTAAGAGCAGAAATTAAT 7479
Db 1681 -----IleGlyTyrGluMetTyrArgAsnLeuAlaGlnGlyArgAsnVal 1695
QY 7480 ATGATATTAATCTTGAAGAGAAATTTATATCTATTTAGTCTACAAAGAAAGAAAGT 7539
Db 1696 LysSerArgLysLeuLysGluIlePhe----- 1704
QY 7540 AAAATATGTTGGAATTAATATTCCTTGGACTTAAGAAACCTTATGCACTGATTA 7599
Db 1705 -----AsnLysAlaLeuValAspPro-----GlyProAsp--- 1714
QY 7600 TATATGAGAGAGAAACCTTGTGAATAATAGAGAAATGCTTTTAAGATGATTA 7659
Db 1715 -----PheValValCysAsp 1719
QY 7660 GAATGGAATGTACAAAAATTTCAAACTTATCAGAGAAAAAGAGTATGTACT 7719
Db 1720 GluGlyHisIleLeuLysAsnGluAlaSerAlaValSerLysAlaMetAsnSerIleArg 1739
QY 7720 CCAAGAA-----GAAATATG 7737
Db 1740 SerArgArgArgIleIleLeuThrGlyThrProLeuGlnAsnAsnLeuIleGlyLysHis 1759
QY 7738 TGCCTTAAGCAATTTAGATCAATTAATTAAGAAAGCTTAAGATGATTAATTTACTCTTA 7797
Db 1760 CysMetValAsnPhe-----IleLysGluAsn-----LeuLeu 1770
QY 7798 AAAATGCTTCGCAACTGCAGCAATGAAGATTAATTAATTAATTAATTAATTAATTA 7857
Db 1771 GlySerIleLysGlu---PheArgAsnArgPheIleAsnProIleGlnAsn----- 1786
QY 7858 GAGAGCGGTGCGCAATGATCAATGTGAT-----ACTATGAA----- 7899
Db 1787 ---GlyGlnCysAlaAspSerThrMetValAspValArgValLysLysArgAlaHis 1805
QY 7900 TATAGTTTCGCTGATCTGCTGACATAGTATAGAGGAACAGATATGTTACAAATGTGTGT 7959
Db 1806 IleLeuTyrGluMetLeuAlaGlyCysValGlnArgLysAspTyrThrAlaLeuThrLys 1825

```

QY 7960 TACTTACCTCC-----GTAGAAATAA 7983.
   :|||:
Db 1826 PheLeuProPolysHisGlyTyrValLeuAlaValArgMetThrSerIleGlnCysLys 1845
QY 7994 TTATATAGGTTTT---GAATACATATATGAAATGAGAAATATAAATAAAGGTGA 8040
   :|||:
Db 1846 LeuYrGlnTyrTyrLeuAspHisLeuThrGlyValGlyAsnAspSerGluGlyArg 1865
QY 8041 AATAAA-----TACAAAGATGTACAAACGTTTCGTTCTGGTGG----- 8079
   :|||:
Db 1866 GlyLysAlaGlyAlaValLeuPheGlnAspPheGlnMetLeuSerArgIleThrPheHis 1885
QY 8080 ---TGGATGCTAATAGAAAAGATTTTGAAGCAATAGCAGTGCAGAACAGAGAT 8136
   :|||:
Db 1886 ProTrpCysLeuGlnLeuAspTyrIleSerLys-----GluAsn 1898
QY 8137 GCAAAACCTTTTAGAAAAGAAAGATGATGATTTGAACGCATATATATACAAAT 8196
   :|||:
Db 1899 LysGlyTyrPheAspGluAspSerMetAspGluPhe-----IleAlaSerAsp 1914
QY 8197 AAGTGTGACATTAAGCAGATCCACCTGTGATGATTAATACCTCAACGGTTTCGATGG 8256
   :|||:
Db 1915 SerAspGluThrSerMetSerLeuSerSerAspAspTyrThrLysLysLys----- 1931
QY 8257 ATGACTGAATGCTGAATATATTATTGTAAGCAGATGGAAGAAATGGAATAATTAAA 8316
   :|||:
Db 1932 -----LysLysGlyLysLysGlyLys 1938
QY 8317 AAATCATGTGATCACTGTAAACATCTGCAGATGCAGAAATGAT----- 8361
   :|||:
Db 1939 Lys-----AspSerSerSerSerLysSerGlySerAspAspAspValGluValIleLys 1956
QY 8362 -----TATGATGAAAT----- 8373
Db 1957 ValTrpAsnSerArgSerArgGlyGlyGlyGluGlyAsnValAspGluThrGlyAsnAsn 1976
QY 8374 -----AAGTGCACAGCTGTAAACGACAGATGCACAA----- 8406
Db 1977 ProSerValSerLeuLysLeuGlnLysSerLysAlaThrSerSerSerAsnProSerSer 1996
QY 8407 -----TATAAAATTTGTTTAAATGGAATCTCTATTCGATATACA 8451
Db 1997 ProAlaProAspTrpTyrIleLysAspPheValIleThrAspAlaAspAlaGluValLeuGlnHis 2016
QY 8452 TCAAATTAATACAAAGAAATGTATGAA-----CAACCAATATATACAAA 8496
   :|||:
Db 2017 SerGlyLysMetValLeuLeuPheGlnIleLeuArgMetAlaGluIleGlyAspLys 2036
QY 8497 ATCTCTACTAT-----GATCATGTTCAAAAATTTGTACAAAAG 8535
   :|||:
Db 2037 ValLeuValPheSerGlnSerLeuIleSerLeuAspLeuIleGluAspPheLeuGlnLeu 2056
QY 8536 TTGAACACTTTTAAAGTGAATGTTCTGTGAGACTTTTGAATATCTCATGAACA 8595
   :|||:
Db 2057 AlaSerArgGluLysThrGlu-----AspLys 2065
QY 8596 ACTAAGTGTGAATTAATTAATTAATGAATAATGATGTTCCATATACGAACATAT 8655
   :|||:
Db 2066 AspLysProLeuIleTyrLys-----GlyGluGlyLysTrpLeuArgAsnIleAspTyrTyr 2084
QY 8656 GCTTCGAAGAAACACAAAAGTTATAAGAGCTGCAGCTTGACACTACTCTCTAAG 8715
   :|||:
Db 2085 Arg----- 2085
QY 8716 AATCCATGATTAATGCTCCTACGATCAAAAC-----AAAGATGATGTAAGAAATTAACA 8772
   :|||:
Db 2086 -----LeuAspGlySerThrThrAlaGlnSerArgLysLysTrpAlaGluPheAsn 2103
QY 8773 ACTTTTACTCTCTGCTCGAAGATGATATGATTAATATCTGATTAATGGAACGATAC 8832
   :|||:
Db 2104 AspGluThrAsnValArgGlyArgLeuPheIleIleSerThrLysAlaGlySerLeuGln 2123
QY 8833 CTTGTTCTTAATA----- 8845

```

```

Db 2123 yIleAsnLeuValAlaAlaAsnArgValIleIlePheAspAlaSerTrpAsnProSerTyr 2143
   :|||:
QY 8846 ---GTTGAGATGATTAACAAAGGTGATTTGATTCCTCCAAAGAAAGACATTTATGA-- 8899
   :|||:
Db 2143 RasPleGlnSerIlePheArgValIleTyrArgPheGlnIleThrLysProValIleValTyr 2163
QY 8900 -----CA 8901
Db 2163 fArgPheLeuAlaGlnGlyThrMetGluAspLysIleTyrAspArgGlnValThrLysGln 2183
QY 8902 AGACCTATCACTGCATATATTAATAGAAAAGCTGATTAAGAAATTTA----- 8949
Db 2183 nSerLeuSerPheArgValValAspGlnGlnGlnValGlnArgHisPheThrMetAsnGln 2203
QY 8950 -----AA 8951
Db 2203 uLeuThrGluLeuTyrThrPheGluProAspLeuLeuAspAspProAsnSerGlnLysLys 2223
QY 8952 AAAAAAATCTTCTTACTTCTGCTTCAGTCAAGCAAAATTTAGTCA----- 9000
   :|||:
Db 2223 sLysLysArgAspThrProMetLeuProLysAspThrIleLeuAlaGluLeuGlnIle 2243
QY 9001 -----AAATATAATGGA 9014
Db 2243 eHisLysGlnHisIleValGlyTyrHisGlnHisAspSerLeuLeuAspHisLysGluGln 2263
QY 9015 AGAAGAGTTGCTTTGAGGCAGAAATATATAGTTATGACATTTATTC----- 9063
   :|||:
Db 2263 uGlnGluLeuThrGlnGluGluArgGlyAlaIleTrpAlaGluIleValGluIleLysLys 2283
QY 9064 -----GATATATTAATAAGCAACTATATGATGCACACTTCATATTC 9104
   :|||:
Db 2283 sValLeuThrMetArgPheAsnIleProThrGlyThrAsnLeuProValSerPheAs 2303
QY 9105 TGAATAAAT----- 9114
   :|||:
Db 2303 nSerGlnThrProTyrIleProPheAsnLeuGlyAlaLeuSerAlaMetSerAsnGlnGln 2323
QY 9115 -----AAAAATATTTGAAACATCAAAATGAAGCAAC 9146
Db 2323 nLeuGluAspLeuIleAsnGlnGlyArgGluLysValValGluAlaThrAsnSerValThr 2343
QY 9147 CGAAATATCT-----AAAAATGCTGGGAAAT 9174
   :|||:
Db 2343 fAlaValArgIleGlnProLeuGluAspIleIleSerAlaValTrpLysGluAsn 2361

```

Search completed: April 28, 2003, 11:32:25
 Job time : 803 secs

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 28, 2003, 11:02:12 ; Search time 296 Seconds

(without alignments)
6903.491 Million cell updates/sec

Title: US-10-087-013-1

Perfect score: 19455
Sequence: 1 atgggggtctctctgcaata.....atatatgatgaagatattg 10628

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame+ n2p model -DEV-xlp
-O-/cgn2.1/USPTO.spool/US10087013.rpr.rpt.28042003.102846.20277/app.query.fasta.1.10823
-Db-pir.73 -OFMT-fasta -SUFFIX-n2p -rpt -MINMATCH-0.1 -IOOPT-0
-UNITS-bits -START-1 -END-1 -MATRIX-blosum62 -TRANS-human40.cdd -LIST-45
-DOCALLGN-200 -THR-SCORE-pct -THR-MAX-100 -THR-MIN-0 -ALIGN-15 -MODE-LOCAL
-USER=US10087013.qcgn.1.1.520.etrnat.28042003.102846.20277 -NCPU-6 -ICPU-3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES-0 -WAIT -LONGLOG -DEV-TIMEOUT-120
-WARN_TIMEOUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5 -FGAPOP-6 -FGAPEXT-7
-YGAPOP-10 -YGAPEXT-0.5 -DELDP-6 -DELEXT-7

Database:

1: PIR_73:.*
2: PIR1:.*
3: PIR2:.*
4: PIR3:.*
5: PIR4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4515.5	23.2	2706	2	T28155	variant-specific s
2	4260	21.9	3006	2	T28625	variant-specific s
3	4010	20.6	3026	2	T28431	variant surface pr
4	3501	18.0	3078	2	T28432	variant-specific s
5	3154	16.2	2924	2	T18378	variant-specific s
6	2727	14.0	2647	2	T28161	hypothetical prote
7	2726.5	14.0	2654	2	T28626	variant-specific s
8	2678.5	13.8	2212	2	T28157	erythrocyte membra
9	2303	11.8	2135	2	T14602	variant-specific s
10	1773	9.1	2197	2	B71600	variant-specific s
11	1755	9.0	2182	2	T28634	variant-specific s
12	1731.5	8.9	2228	2	T14029	variant-specific s
13	1586	8.2	1711	2	C71625	variant-specific s
14	1542.5	7.9	2042	2	T18399	variant-specific s

15	1249	6.4	1729	2	T18396	erythrocyte membra
16	835.5	4.3	431	2	F71600	variant-specific s
17	618	3.2	1435	2	A37793	erythrocyte-bindin
18	561.5	2.9	4550	2	T18440	hypothetical prote
19	558	2.9	3724	2	T18427	hypothetical prote
20	539.5	2.8	440	2	E71625	variant-specific s
21	535.5	2.8	3394	2	T18501	hypothetical prote
22	513.5	2.6	4961	2	T18489	hypothetical prote
23	461	2.3	2523	2	T18402	hypothetical prote
24	453.5	2.3	3844	2	T18402	asparagine/asparta
25	450.5	2.3	3973	2	B71612	hypothetical prote
26	443.5	2.3	2401	2	T28676	hypothetical prote
27	435	2.2	2829	2	A42771	rhodopsin
28	434	2.2	2510	2	T28160	erythrocyte-bindi
29	427.5	2.2	2657	2	T18497	hypothetical prote
30	427	2.2	1979	2	C71622	hypothetical prote
31	424.5	2.2	2437	2	T18482	hypothetical prote
32	420.5	2.2	6669	2	S55024	nebulin, skeletal
33	409	2.2	2269	2	T28677	rhodopsin
34	409	2.1	2485	2	C71624	variant-specific s
35	404	2.1	2573	2	D71614	hypothetical prote
36	392	2.0	2485	1	H71621	serine/threonine-s
37	387	2.0	4688	2	F82885	hypothetical prote
38	384.5	2.0	3135	2	A48584	transmission block
39	366.5	1.9	1939	2	T18372	repeat organellar
40	366	1.9	2295	2	B71621	probable membrane
41	362	1.9	1819	2	A71928	cag island protein
42	362	1.9	2539	2	B71619	hypothetical prote
43	357.5	1.8	2500	2	G71609	hypothetical prote
44	355	1.8	1927	2	G64585	cag pathogenicity
45	354.5	1.8	1045	2	T18373	erythrocyte bindin

ALIGNMENTS

RESULT 1

T28155 variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum) (fragme
N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)
C:Species: Plasmodium falciparum

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 20-Jun-2000

C/Accession: T28155
R/Rowe, J.A.; Moulds, J.M.; Newbold, C.I.; Miller, L.H.

Nature 388, 292-295, 1997

A>Title: Plasmodium falciparum rosetting is mediated by PfEMP1 and requires complemen

A/Reference number: Z20477; PMID:97373957; PMID:9230440

A/Accession: T28155

A>Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-2706 <ROW>

A/Cross-references: EMBL:Y13402; PDB:CAA73831.1

A/Experimental source: strain IT 4/25/5

C/Genetics:

A/Introns: 2493/3

A/Note: R29R+var1

Alignment Scores:				
Pred. No.:	1.19e-214	Length:	2706	
Score:	4515.50	Matches:	1191	
Percent Similarity:	41.69%	Conservative:	372	
Best Local Similarity:	31.77%	Mismatches:	871	
Query Match:	23.21%	Indels:	1315	
DB:	2	Gaps:	123	
US-10-087-013-1 (1-10628) x T28155 (1-2706)				
OY	100	AGTCAAGAAATGTTTGGACGTTATGCCAAAAATTAAGACATCCATCAAAA-----	153	
DB	14	Serialatraspvalleuglutylsilealathrcilyleryrsnginglutylsglutys	33	
OY	154	---TATCAAAAGAAATGCGTGAAGAGGATTTGAGCAAGAAAGAAATG---207		
DB	34	ValtyrProtyrGlu-----AsnGluLeuylsGlylleuSerAsnAlaIlePheVal	51	

```
QY 208 -----CGTGGTCTCTTCTACGCCAGTAATAG 237
Db 52 AspGlnLeuArgLysGlnLeuAsn1IegLuserProGlyProSerAsp----- 67
QY 238 CATATAATATATATATCATATCATATTTTGTATATAGACATACTATATTTACGG 297
Db 68 -----SerCysSerLeuAspHisIstPheHisIleAsn 80
QY 298 TATGATGATGTGAATTTAGACATCTTGGCATGTGTAGACAAAACGATTTGTATGA 357
Db 81 ThrGlnThrGlnGlnArgLysProCysTyrGlnArgGlnGlnLysArgPheSerAsn 100
QY 358 GATGAGAAATCTGAATGTGA---AATAAATACGTATATTAAGAAAATATGTCT 414
Db 101 GlnGlnGlnAlaLysCysGlnSerAspLysIleArgAspArgLysIleLysSerAlaGly 120
QY 415 ATAGCTGTGCGCCACCTAGACAGCATGTGTGTATTAAGAAAATGTGAAGCTGTAAAT 474
Db 121 GlnAlaCysAlaLeuProPheArgGlnAsnLeuCysAspArgAsnLeuGlnLysIle 140
QY 475 GATATATAATACCCAAAATATTCATGATTTATTTGGAAATGTACTAGTACAGCAAAATAC 534
Db 141 AsnLysAsnThrAsnThrThrHisAspLeuGlnLysValLeuValThrAlaLysTyr 160
QY 535 GAAGGTGAATCATGTGTATATATCATCCATCAATAAAGAACT-----TCAGAC 582
Db 161 GlnGlnLysSerIleValAsnAsnHisProAspLysAsnSerSerGlyAsnLysSerSer 180
QY 583 GCTTGTACTGCTCTTGCACAGATTTTGCAGATATAGTGATATTTAGAGAGAAATAG 642
Db 181 IleCysThrAlaLeuAlaArgSerPheAlaAspIleGlyAspIleValArgGlnArgAsp 200
QY 643 ATGTTTAAACCAAAATCTCATGACAAAGTAGAAACGGCTCCGAGAGTTTTCAGAAA 702
Db 201 MetPheLysProAsnAspAlaAspLysValGlnLysGlnValPheGlnLys 220
QY 703 ATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 759
Db 221 IleTyrAsnSerLeuProSerProAlaGlnLysHisTyrAlaHisAspArgLysSerGly 240
QY 760 AATTTATTAATTAATTAAGACAGACGTGGTGAATGTGAATGAATGAATGAATGAATGA 819
Db 241 AsnTyrTyrLysLeuArgLysArgLysArgLysArgLysArgLysArgLysArgLys 260
QY 820 ATAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 879
Db 261 IleThrCysArgAlaProAsnGlnLysAlaAsnPhePheArgAsnIleSerGlyAsnMetLys 280
QY 880 TTATTTT---TCAAAATCTTAATATGCGCCCATTAACAGAAAGATGCTTACCAATTTAGAT 936
Db 281 AlaPheThrSerGlnGlyTyrCysGlnHisSerGlnLysHisValProThrAsnLeuAsp 300
QY 937 TATGTCCCTCATATTTACGTGGTGGTCCAGCAATGGGAGAGAGATTTGCCGAAAAA 996
Db 301 TTTTATTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1016
QY 997 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1056
Db 321 LysIleLysLeuGlnLysValLysLysLysLysLysLysLysLysLysLysLysLys 340
QY 1057 AGTCATTAATGACATGATGTGTACGACAACTATTTGGAAAAAAGATTTTGTGATTTGAT 1116
Db 341 SerGlnAspArgLysHisAspCysLysArgLysLysLysLysLysLysLysLysLys 360
QY 1117 AATAAGTCTACTGACTGTTCGACTAATGCAAAAGTTTGTGAAGTTTGGTTAGGCAATCA 1176
Db 361 LeuAsnCysProArgCysGlnLysAlaCysSerAsnTyrLysTyrLysIleGln 380
QY 1177 CAAGAGAGATTTTAAACCAAAAGAAAGAAATATGAAAAAGAAATATATTTATACG 1236
Db 381 ArgLysGlnPheAspLysGlnLysArgLysTyrMetAsnGlnIle----- 395

QY 1237 AACGATAACAAATTTTGCATATAATTT---AATAGGATATATTAACATTTTATGAA 1293
Db 396 -----LysIleLysThrAsnIleSerAsnAspGlnAsnAspLysGlnPheTyrGln 412
QY 1294 AAACCTTAAGAAAACCGCAATATGCAACTTATTAATTTTAAATTTTAAATGAAGA 1353
Db 413 AsnLeuAspLysLysGlnLysSerThrIleAsnThrPheLysGlnSerLeuAsnHisGly 432
QY 1354 AAGTATTTGAAGAGATTTACAGAGAAAAGATATTACTTTTACTAATGAGTCTGTAT 1413
Db 433 LysGlnCysGlnHisAspAsnIleAspLysLysAsnLysThrAsnPheLysAsnLeuGln 452
QY 1414 GACAAAGGATATTTTATGCTCAGAAATGTGCAAGTGTGCGGACGTCCGAGGTCGAAA 1473
Db 453 -----ThrPheLysProSerGlyTyrCysGlnAlaCysProIleTyrGlnValLys 469
QY 1474 TGTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1533
Db 470 CysSerAsnGlnLysCysThrProValThrGlnAsnGlnLysPhe-----AsnSer 485
QY 1534 GACTATTAACCTCCATGCG-----GGTGTGAAGCTTACTAATATCTACTGTC 1578
Db 486 AsnAsnAlaGlnLeuProThrAspThrSerThrLysAsnLeuAsnAlaThrAsnIleAspMet 505
QY 1579 CTTTATAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1638
Db 506 LeuValAsnAspGlyIleGlyLysAlaIleAspAsnGlnLysLysAsnCysThrLys 525
QY 1639 TCACTAATTTCAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1698
Db 526 TyrGlnLysIleLeuLysGlyIleLysLysGlnLysTyrGlnCysGlnTyrLysLeuAsn---Asn 544
QY 1699 ATTAATAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1749
Db 545 IleAspGlnCysLysIle-----AsnAsnValLeuAsnSerGlyTyrPheAspAsnLys 562
QY 1750 AAGATATATCATCTTCAATTAATTTTGTGAATTAATGCGTTTACATTAATTAATTAAT 1809
Db 563 -----IleAlaPheAsnValLeuPheGlnArgTyrPheValArgAspHis 580
QY 1810 ATTAAGTGAATGACAAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1860
Db 581 AsnArgLysGlnLysIleAspValCysIleLysLysGlnAsnIleAsnGlnAsnIle 600
QY 1861 TGTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
Db 601 CysIleLysArgCysLysThrAsnCysGlnLysValGlyLysTyrPheLysLysGln 620
QY 1921 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1971
Db 621 AlaGlnTyrAspLysIleAsnGlnHisTyrAsnGlnLysAsnHisIleMetPheIleLeu 640
QY 1972 -----CAATGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2016
Db 641 IleProTyrTyrIleThrGlnGlyPheTyrGlnLysIleThr-----PheProAsnAspPhe 638
QY 2017 TTTTAAAGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2076
Db 659 PheLysAlaLeuGlnLysValAspThrIleAsn----- 669
QY 2077 ATTAAGAAAGAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2136
Db 669 ----- 669
QY 2137 GCATTAAGACTCTTGTAGATCACTTAAGAAAGAACTGCGACGATATGTAAGCAATTAAT 2196
Db 670 -----ValLeuAspThrIleLysGln----- 676
QY 2197 ACAAGACAGACATGCTTCAATCAATGACAAACAAACCCGGTGTAAACCTCGT 2256
Db 677 -----CysGlnAspThrHis----- 681
QY 2257 GGAAGCAGCAACCACTAATAAATATTAAGAAATTAATGACCAATCACTTTAAAGAGATGCA 2316
```

[illegible]

Dd	740	----	AspGIuGIuLyGIu-----	744
Qy	3397	ATGCTGTAGTAATAGTGGTATTGAAGCTTCCAGTACTCCAAAAATCATATTACACAGG	3456	
Dd	745	-----	GIuSerSerGIuThrThLyGArGAsnIle-----	754
Qy	3457	AATGTATTGAATTTTGTGCGAATTATACCAACAAAAATGGTGGCAAAATGATTAAGT	3516	
Dd	755	-----	SerGIuLySGLy 758	
Qy	3517	GGTACTAGTATGAAGTCTGTCAATGGTACTAACACCACTATGAAATGTGGACCA	3576	
Dd	759	GIYThrLySserAlaSerCySValLySGLyAlaCySAlaIleValLySGLyAla---	777	
Qy	3577	TATCTCCATGATACAGAAATTTTGGATGTGCATGCTCACTCAACAAATGAGTTTGTGATGA	3636	
Dd	778	GIuGIuLySserAsnGIySerIleAspAsnCysAsnAlaLySAsnArg-----	793	
Qy	3637	AAAAGTGATGTAAAGATTAACGAAAAATATGCTTTAGAGATTAACACAGGACCATCAT	3696	
Dd	793	-----	793	
Qy	3697	GGTGGCTGTGGTTTAAAGTGATGCAAAACGACAGGGGTACAGATTAACAAAAA	3766	
Dd	793	-----	793	
Qy	3757	AAAGCGAAGAAAGAGATACGGAATGTAAACAGTGAATGATATACTTAAAGAAAAAGAT	3816	
Dd	793	-----	793	
Qy	3817	GGAAAAAACAAGTAGAAGATTGTCAATCCAAAAAGAAATAGTAATGATATCCGATTGG	3876	
Dd	794	-----	LyLySAsn-----	
Qy	3877	CAATGGCGGAATATTAATATTAGTGAA---GACCCCTGGTGGTATGCCCTTCGAAGA	3933	
Dd	799	GIuCySAspLySAsnThrPheValAspGLyAsnGIuLyAlaCySAsnProForCAtArg	818	
Qy	3934	CAAAAGTTATGCGTATTTCTTGGCAATGATTAATGAATTAATAAATTTACAAATCAAA	3993	
Dd	819	LySserIleCySAlaIleAsnLeuThrLeuGIuGIuGIuThrLyS-----	835	
Qy	3994	GTTATTTAAAAAGAGTTTCATCAAAATCGACAGCAGCAAAATCTTCTCATAGCAT	4053	
Dd	836	TyrGIuLeuATgGIuAlaPheIleLySAsnAlaLySGLuThrAsnLeuLySAsp	855	
Qy	4054	TATTATTAAGTAAAGATGTGTGAAGAAATGAAGTCAATGAATTAAGAAAGACGACAA	4113	
Dd	856	LySThrLySAsnAspLySAsnGIuAlaGIuLeu-----	872	
Qy	4114	ATTCCTCCCGCATTTTGAAGATCCAGTCTTACACAATTTGGAGATTAAGACATTTTAA	4173	
Dd	873	IleProGIuAspPheMetArgIleMetPheTyrThrPheGIyAspPheArgAspPheCyS	892	
Qy	4174	TTTGGAACAGATTAATCAAAAAGGTCAGTGGAGGAAGTAACATAAAGACGATTAAT	4233	
Dd	893	LeuGIuAsnAspMetGIyLyS-----	909	
Qy	4234	TCTCTTTCAAAATGGTGCACAAAATCTCTTAATGAAAAAACAAGCCCAAGATGTGG	4293	
Dd	910	LySValPheAsnAsnSerSerLySArgLyPheLyLyIleAspProGIuAsnThrPir	929	
Qy	4294	ACAGAACATTAATCATGATATGAGTGAAGCATATCTATGTGCATCTAGTAATAAATTTGGGCA	4353	
Dd	930	AsnGIuAsnGIyProGIuIleTrrPAsnGIyMetLeuCySAlaLeuIleHisAlaAspThr	949	
Qy	4354	AAA-----	4398	
Dd	950	LySAspSerIleLySAsnLySAsp-----	965	
Qy	4399	TTTACTGACAAA-----	4443	
Dd	966	IleLeuAlaLySAspGIySerAsnGIyMetThrLeuSerGIuPheAlaLySAspPro	985	

Db 1676 IleLysLysIleGlnGlnIutPalaSerGlyAsnThrGlnIaGlyLysAsnThrGlySer 1695
 QY 6442 -----AATAGCGTGCACATGATATATA----- 6465
 Db 1696 GlyAspSerThrGlyAlaSerSerAsnProGlnGlyAsnGlyValSerThrSerProGln 1715
 QY 6466 -----GAGCGTTTATTAATGTCAGCAATAGCAATCAATTTGTGGTTA 6513
 Db 1716 ValAlaLeuLeuHisAlaPheValGlySerAlaIaValGlnIutThrPheLeuThrPasp 1735
 QY 6514 AATAT----- 6528
 Db 1736 ArgTyrLysLysGlyLysGlyIleGlyLysGlnGlnGlnIutThrGlyLeuValAla 1755
 QY 6529 AATCGTCGACGAGA-----AATGATGCAAAATGAGCAATTCGA 6570
 Db 1756 SerGlnIutSerGlnAspThrGlnHisProGlnAsnLysLeuGlnSerGlyThrIlePro 1775
 QY 6571 GATGAATTTAAAGCAATATGATATATACATATGCGTATATTAAGATATGTTTGTGA 6630
 Db 1776 LeuAspPheLeuArgGlnMetPheTyrThrLeuGlyAspTyrArgAspIleCysValGly 1795
 QY 6631 ACGATATTTCTATGATATAAAAAATATATACCTGAACAATAGCTGAACAACCATCTC 6690
 Db 1796 -----LysThrProAspGlyIleAspThrValSerIaSerAspLysAspThrMet 1812
 QY 6691 AATGAAAATTAATAAGAAAAACAGAT-----AAAAAAAAGAT----- 6729
 Db 1813 AspLysIleGlnIaLysIleGlnGlnIleLeuProLysLysAspThrProProSerVal 1832
 QY 6730 GAAGAAATTAAGCAATATTTGTGGAGAAAAATTAATTTGTGGAGAAATGATA 6789
 Db 1833 LysThrProGlnIutThrPaspLysHisIaGlnIutThrPaspGlnIle 1852
 QY 6790 TATGATTTACTATGATCTCCACAGCAAGAAAAAGAAAAATTAATA----- 6840
 Db 1853 TyrAlaLeuThrTyrLys-----ThrAspThrProSerGlyLysProLysGlnIlePro 1871
 QY 6841 -----GATATTTACAGTAC 6855
 Db 1872 GluValLysThrLysLeuPheAspGlyLysGlyThrProGlnSerAsnLysTyrGlnIut 1891
 QY 6856 AATGACATGACCAACTG----- 6873
 Db 1892 LysThrVal-----LysLeuGlnGlnIutThrSerGlyAlaLysProLysSerThrGlu 1910
 QY 6874 -----ACGCGCTCC-----CTTGAGAGCTTTGTAAAAAGCCCAATTT 6912
 Db 1911 SerSerSerProSerGlyLysAsnThrProLeuThrAspPheIleSerArgProTyr 1930
 QY 6913 TTGAGATGTTTCACAGAAATGGCAGAAATTTGTAAATGAAGAAAGAAAGTGTGA 6972
 Db 1931 PheArgTyrLeuGlnIutThrPaspGlyLysAsnThrPheCysLysGlnIutThrLysAspGly 1950
 QY 6973 AAATTTGAGCGCGCTGT-----AAGCAATATGAGTCTAATGCT----- 7011
 Db 1951 AspIleLysSerAsnLysLeuLysAspGlyAspLysGlnIutThrSerGlyAspGlyAla 1970
 QY 7012 -----AGTAT-----GACGGTATGACACAAAGATGT 7038
 Db 1971 CysSerAsnIleAspValAsnLysAspLysIlePheAlaAspLeuGlnIutProLysCys 1990
 QY 7039 GCAGAGCGGTGTGTACATATTAATTTATTAAGAAAGTGAAGAAAGTGAATGAAGA 7098
 Db 1991 AlaLysProCysSerSerTyrArgLysTyrIleLysGlyLysLysThrGlnIutThrGlyLys 2010
 QY 7099 CAAGAGAAAAAGTCAAAAAAGATTAAGATGCAAAAAAGATTAAG-----GATTAAT 7149
 Db 2011 GlnIutLysLysLysGlnIutThrLysGlnIutCysLysLysGlnIutThrAsnAsnHisTyr 2030
 QY 7150 CTTTCTATGAAAGACATAGAGAAAGCAACATGCTCATGATTTTAACATGAAG 7209

Db 2031 AsnGlyPheArgValLysLeuGlyThrCysThrThrAlaGlyAspPheLeu----- 2047
 QY 7210 TTAAGAAATTTATGCGCATTAAGATTTGTTGTATGCAAAAACCTTCTCAACAATA 7269
 Db 2048 -----GlnThrLeuLysAsnGlyProCys----- 2055
 QY 7270 CCAAAAAACAACACACAAATCACAATCCGATTCGATATATATGCGAGAAATCGTGAT 7329
 Db 2056 -----LysSerGlnAsnGlyLysAspHisGlnIut----- 2065
 QY 7330 TATGTTCTGAAAGATTTTAACAAGTGTGAGTGTCCGAACTTTCAAAAAAGGATCTATG 7389
 Db 2066 -----AspIutIleAsnPheSerGln-----ProAspValIaThrPheArgProAla--- 2080
 QY 7390 ATTCAATCAAAAAAATTAATGAAACCTTAATAATACCTATGATTTGTGAAGAAAGCA 7449
 Db 2081 -----ThrAsnCysThrProCysProLysPheLysValAsnLys----- 2093
 QY 7450 TATTAATTTATCTAAGACGAGAAATATATATGATATATACCTGAAGAAAAATTTATA 7509
 Db 2093 ----- 2093
 QY 7510 CTTATGAGTCTCAAAAAGAAAGAAAGTAAAAATAGTTGACATATATATCTTGC 7569
 Db 2094 -----LysAsnGlyAsnCysGlyAlaAspThrAsnGlyLysCys 2106
 QY 7570 GATCTTAAGAAACCTTATGCACTGATTAATATATAGGAAGAAACCTTGTGAANAAT 7629
 Db 2107 AsnGlyLysThrPro----- 2111
 QY 7630 AGAGAAAGAAATCGTTTAAGTAGATTAATGAATGAATGTACAAAAATTCAAAAGTTC 7689
 Db 2111 ----- 2111
 QY 7690 TATCAGAGAAAAAAGATGATGTCTACCTCAAGAAAGAAACATATATGCTTAAGAAAT 7749
 Db 2111 ----- 2111
 QY 7750 TTAGATGAATTAATTAATTTGAAGACTTAAGATAGTAAATATCTTAATATGTTTCGT 7809
 Db 2112 IleAspAlaGlnAsnPheGlnGlnMetGlyGlnThrAlaLysGlnPheValIleLeuVal 2131
 QY 7810 CGAATGTCAGCAATGAGAAATGACATATATTAATAAAATCTCAACTCAGAGAAAGCGGTGC 7869
 Db 2132 SerAspLysSerThrAsnGlyPheGlnVal-----AsnAspLeuAsnGlnLys 2147
 QY 7870 GCAATGAATCCATATGTATATGATATGATATATATATATATGCTGATCGGTGACATATGT 7929
 Db 2148 Ala----- 2148
 QY 7930 AGAGAAACAGATATGTTACGAATTTGTTACTTACCTCCGTAGAAATTAATTAATAT 7989
 Db 2149 -----GlyAlaAsp----- 2151
 QY 7990 AAGCTTTTGAATACATATATGCAAAATGAGAAATTAATAAGGTAGAAATTAATAC 8049
 Db 2152 -----IlePheGlnGlyIle----- 2156
 QY 8050 AACGATGTACAAACGTTCTGCTTGTGTTGGATGCTAATTAAGAAAGATTTGGAAA 8109
 Db 2157 -----LysGlnAsnIleThrPasp 2162
 QY 8110 GCAATG-----ACGTGCAAAAGCACAGAA-----GATGCAAAA 8142
 Db 2163 CysArgAsnValCysGlyLeuAspValCysLys-----ProGlnLysValAsnAspGlnLys 2181
 QY 8143 CTTTATTTGAAGAAGAGAAATGATGATTTGAACGCAATTAATTAATCAAGATAGTGT 8202
 Db 2182 ValAsnGlyLys-----GluAsnAspGly----- 2189
 QY 8203 GGAATTAAGACGATCACTGTTGATGATTAATATACCAAGCGTTTCGATGATGATGAT 8262
 Db 2190 -----ThrTyrIleIleGlnIleArgAlaLeuLeuArg 2200

Qy	8263	GAATGCTGCATATTATTGTAAGCACTGATGGAAGAAATTTGAAAAATCA	8322
Db	2201	ArgTPrValGIlaAspHe	2216
Qy	8323	TGTGATCACTGTAAACATCTGCAGATGCGAAGAAATGATTATGATAAATAGTGCAA	8382
Db	2217	IleSerHisCysThrAsnSerThrGIuGLuLys	2232
Qy	8383	CAGGTGAACAGAGATGTGCAGAAATATAAAAATTTGTCTTAATGAAATCTGTATTC	8442
Db	2233	---CysGLyLysLysCysLys---	2247
Qy	8443	GATATACATCAATCAAAATATCAAAAGAAATGTATGAACAACCATATATATACAAATCTCT	8502
Db	2248	LysGIuGLuTrpLysLysIleLysGIuHisTyr	2258
Qy	8503	ACTTATGATCATGTTCAAAATTTGTACAAAGCTGTAAACCTTTAAAGTGAAATGTTCT	8562
Db	2259	-----ValGIuLysAsnLysSerGIu-----	2265
Qy	8563	GTTGAGACCTTTTCGAAATATCTTCATGAAACAAGTAAGTGTGATTAATTAATTAAT	8622
Db	2266	---AspAspPheSerAspAsnIleAsnSerPheLeuGIuThrLeuIleThrGIuLLeuPro	2284
Qy	8623	GAAATGATGCTTCTCCAAATATACGAACATATGCTTTGGAAGAAACACCAAAAAATTAT	8682
Db	2285	ValIaAspValGIuGIuAsnValIleLysLeuSerAsnPheAspThrPro	2301
Qy	8683	AAAGAGCTTCAGTATGTACACACTTCCTACAGAAATCCATTGGATTAATGTTCACCGAT	8742
Db	2302	-----CysGLyCysSerIaAspAlaAsnSer---	2310
Qy	8743	CAAAACAAGATGAGATGTAAAGATTACAAACCTTTACCTCTGCTCGAAGATGATTAT	8802
Db	2311	GIuLysLysAspGLy-----AsnGIuAsnAspAla	2320
Qy	8803	GATATATCTTGATTAATTTGGAACGCATCTTGTTTAATAGTTCAGATGATACAAA	8862
Db	2321	IleAspCysMetIleAsn-----ArgLeuGIuGIuLysAlaLys	2333
Qy	8863	GGGTATGATTCCTCCAGAAAGAACATTTATGTACAAAGCATTCATGCATTAAT	8922
Db	2334	AsnCysGIuThrGIuProSerGIySerLysGIuCysThrProProThrLeu	2352
Qy	8923	TATGAAAAAGTGATTAAGAAATTTAAAAAAAACCTTCTACTCTGCTTCAGTCAA	8982
Db	2352	-----	2352
Qy	8983	GGACAATTTAGTGCAAAAATTAATTAATGGAAGAAGAGTTGCTTTGAGGCAATGAA	9042
Db	2353	---GluAspPheSerAspAsnLeuThrPheAsp---	2359
Qy	9043	TATAGTTATGCAGATTATTCGCATTAATTAAGAACTGATATGAGCACTTCATTA	9102
Db	2360	-----AspAspIle---	2362
Qy	9103	TCTGAAAAATTAATAAAATATTGTAACATCAATGGAAGAACCGAAATTCGTAACA	9162
Db	2363	-----GluThrAspAsn---	2366
Qy	9163	TGTGTGAAAAATATAGAGCTGAGATATGCGACGGTATGTATGTGATTAATTTGCT	9222
Db	2367	-----ProValIaHisProGIuIleCysGIuGLuValIleGIu	2379
Qy	9223	ACTTCAAAAGTAAATCAATTAATAGTAAGATGCTGTCAATTAACCAAGATGAAGAACTAAT	9282
Db	2380	ThrGIuThrGIuThrValLeuGIuLysCysValaIaProProlaGIuGIuLysLys	2399
Qy	9283	CAGTTCTTCGTTGTTATTAATGAAGGCAAGCAAGCATGTAAAGAAAGAAACATGTA	9342
Db	2400	Asp-----GluLysLysLysPro	2405

QY	9343	AGCATTCATTAATAAACAATAATGTCGCGTTCACAAAGCAAGTAATTTTGAAGCGTCAGAA	9402
Db	2406	GLUAspPro-----	2408
QY	9403	TTATTAAAGACACCTGGATGTCAGATGANTATTAGAAAAATATATATGCTGGAATATATTG	9462
Db	2408	-----	2408
QY	9463	ATAAAAATACAAATGAGAAATCTAAATATATAAATTAAGCAATTAAAAAGATCAATCTTCA	9522
Db	2409	-----	2413
QY	9523	GGTAAATATAGACAATAA-----CCATCTGAGAAAAATGTTCTAGTCATATATA	9570
Db	2414	GLYAlaIleGlyProSerGlyProAlaGlyProGlnGluProSerProThrsSer-----	2431
QY	9571	AAATCAAAAGATTCCTCAATGCGCTTTGGAGTTAAATGATATAAATGAATAGTTACAGGA	9630
Db	2431	-----	2431
QY	9631	ACAAAAATATGAAAAATATGAATTCAGAAAGTACTAAAAAAATATATCCGGTTA	9690
Db	2431	-----	2431
QY	9691	TATTTGTTCAAGATGACAACACAAAAATCATGTACTAGATGGAATATATAAAAAGAA	9750
Db	2432	-----AspAspThrGlnGluAsnHisVal-----	2439
QY	9751	GAGCAAAAGTGTGCTCCTAAAGACGCTATTTCTTAACCCCATGAGATCTTCTTCTAT	9810
Db	2440	-----ThrProGlnLysProProPro-----	2447
QY	9811	CAAGACCTTTATCTTCACACACATCGAGTAGCAAA-----	9846
Db	2448	ProAlaPro-----AspThrArgProValProSerProLeuProProAlaAspGln	2465
QY	9847	---TATGATCCCTAAATATGATATATTGAAAAAGTAGATCTCTGTTGTTATTGTATCGCGG	9903
Db	2466	ProPhaAspPro-----ThrIleuGlnThrThrIleProPheGlyValAlaIleuAla	2483
QY	9904	TTAGTTTGATAGCGCTTCATTCATGAGAAAAAATTCAAATCGTCGAGCACTTGTTG	9963
Db	2484	LeuGlySerIleAlaPhePheMetLysLysLysThrLysHisProValAspLeuPhe	2503
QY	9964	CGATATCTAAATATCCCGCAGAGAGAGTAGTAGATCCCTACGTTGGAATCCAAAAATAG	10023
Db	2504	SerValIleAsnThrProLysGlnAspTyrAspIleProThrIleuLysSerLysAsnArg	2523
QY	10024	TACATACCATTAAGAGTAGTGCTCATCTTAAGCAAAACATATATATATATGGAAGAGAT	10083
Db	2524	TyrIleProTyrArgSerGlySerGlyLysLysThrTyrIleTyrMetGlnLysAsp	2543
QY	10084	ACTAGTGGAGATGAGATATATATATGTGGACCTATCTTCTCTGATATATCTTCATCC	10143
Db	2544	Ser-----AspGlnAspLysTyrAlaPheMetSerAspThrThrAspIleThrsSer	2561
QY	10144	GAAGAGGATGAGGAAGATTTGATATTAATATATATATATATATATATATATATATAT	10203
Db	2562	GlnSerGlnTyrGlnGlnIleuAsnLysIleAsnAspIleTyrValProGlySerProLysTyr	2581
QY	10204	AAAACATGATGAGAGTAGTACTATAACCATCAAAAAGG-----GATATACCA---AGT	10254
Db	2582	LysThrIleuIleGlnValValLeuGlnProSerLysSerAsnGlyAsnIleProHisSer	2601
QY	10255	GATGATACCAACAGTAATGAT---ACACCAGCTACGAATAGTATTATTTGATGATGAATGG	10311
Db	2602	AlaGlyLysProLeuAsnAspMetValProThrThrAsnThrPheThrAspLysGlnTyr	2621
QY	10312	AATGAACGAAACATGATTTTGTTATCTCAATATTTTAAACAAATACGAACCAATATAT---	10368
Db	2622	AsnGlnLeuLysHisAspPheIleSerGlnTyrIleGlnSerArgLeuProMetAspVal	2641
QY	10369	---AATTACAAA---AGTGCAGATATTTCCAAATGAATACGAACCTAAATATCTTATATCT	10422

Db 2642 ProGltTyrAspValSerThrGluLeuProMetAsnIleValGlyAsnValLeu---Asp 2660
10423 GATAATCTCGAGAAAAACCTTTATTAATCATCATAGAGGATTAATACGAGG 10482
2661 AspGlyIleAsnGluSerProPheIleThrSerIleHisAspArgAsnLeuTyrAsnGly 2680
Qy 10483 AAGAAATTAAGTTAAATTAATTAATGAGTACTTAATTAATTAATGATATTCATGAAAT 10542
Db 2681 GluGluIleSerTyrAsnIleAsnMet-----ValAsnAsnAspIleProMetSer 2697
Qy 10543 GCTAGAAATGATCTTATAGAGGTATA 10569
Db 2698 AspLysAsnGlyThrTyrThrGlyIle 2706
RESULT 2
T28625
variant-specific surface protein 3 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
Accession: T28625
Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Gulnet, F.; Herfelde, J.A.; Peterson, D.S.
Cell 82, 89-100, 1995
A:Title: The large diverse gene family var encodes proteins involved in cytoadherence an
A:Reference number: Z20487; MUID:95330813; PMID:7606788
A:Accession: T28625
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3006 <SUX>
A:Cross-references: EMBL:L40609; NID:g886376; PID:g886377; PIDN:AAA75397.1
C:Genetics:
A:Gene: var-3
A:Introns: 2597/3
Alignment Scores:
Pred. No.: 4 286-202 Length: 3006
Score: 4260.00 Matches: 1143
Percent Similarity: 43.848 Conservative: 483
Best Local Similarity: 30.828 Mismatches: 1005
Query Match: 21.908 Indels: 1079
Gaps: 134
US-10-087-013-1 (1-10628) x T28625 (1-3006)
Qy 91 AGTCACAAAGTCAGAAATGTTTGAACGTTATGCCAAAAATTAAGACAT---CCA 147
Db 8 ThrAsnLysThrAlaLeuThrTyrAlaAsnAspLeuGlnGlyAsnValGlnGluIleValAla 27
148 TCAAAATATCCAAAGAACATGTCGATTCGTTCAAGGCGATTTGACGAAGCAGAAATTT 207
Db 28 ThrGluAspAlaLeuThrTyrAlaAsnAspLeuGlnGlyAsnLeuSerGlnAlaLysPhe 47
Qy 208 CGTGGTGTCCTTCTACGCCAGTAATTAAGCATTAATTAATTAATTCATCCATGCAAT 267
Db 48 HisGly-----ValProIleAspValLysAsn-----ProCysAsp 59
Qy 268 TTAGATCATTAAGACATTAATTAATTAAGCATGATGATGATGATGATGATGATGATGAT 327
Db 60 LeuAsnTyrGluIleHisThrAsnVal---ValLysGlyArgLysGluAsnProCys 78
Qy 328 CATGCTAGAGACAAACCCATTGATGAGAGAGACATCTGAATGTCGAAATTAATA 387
Db 79 ArgGlyArgGluIleLysAspPheSerAspValLeuSerGlyGlnSalAlaLysAsnThr 98
Qy 388 CATAATTAATAAGAAAAATGATGCTATGCTGTCGCGACCTAGACAGACATATG 447
Db 99 IleLysAspSerValThrAsnSerValGlyAlaCysAlaProPheArgGluHisVal 118
Qy 448 TGTGATAAAACTGGAAGCTTAATGATTAATTAATTAATTAATTAATTAATTAATTAAT 507
Db 119 CysAspArgAsnLeuGlnIleLysThrAspGluAsnThrSerThrHisAspIleLeu 138
Qy 508 GGAATGCTACTACTTACAGCAAAATAGACAGAGTGAATCAATGTTTAATTAATCAATCCAT 567

Db 139 ValAspValLeuThrAlaLysHisGluGlnSerLeuValLysHisTyrLysGlu 158
Qy 568 -----AAGAACCTCAGACGCTTGACCTCTCTGCACGAATTTTGCAAT 615
Db 159 TyrIleLysLysAsnArgAsnPheAsnIleCysThrValLeuAlaArgSerPheAlaAsp 178
Qy 616 ATAGTGATATTTGTAAGACGAATAGATATGTT-----AAACCAAT 657
Db 179 IleGlyAspIleValArgGlyLysAspLeuTyrLeuGlyTyrAspAspGluIleVal 198
Qy 658 GTCCATGACCAAGTAGAAAGCGGTCCGAGAGGTTTCAAGAAATACATGATGCAATG 717
Db 199 LysArgLysGlnLeuGluLysAsnLeuLysPhePheGlnLysIleHisAspAspVal 218
Qy 718 -----GAAGATGACGTAATAAATGATTAATCAAT 744
Db 219 MetLysThrSerGlyArgThrAsnGlyLysLysSerAlaGluAlaGlnLysArgTyrAsn 238
Qy 745 CCGATGATCTGGAATTAATTAATTAATTAAGACAGACATGTAATGTAATAGAAAT 804
Db 239 ---AspAlaThrGlnLysAsnTyrTyrLysLeuArgGluAspTrpTrpAsnAlaAsnArgAsp 257
Qy 805 AAGATGGAAGACCTATACATGATGATCATATTAATCTGATATTTTATGCAATCA 864
Db 258 GlnValTyrLysAlaIleThrCysAspAlaAlaAspAsnAspGluTyrPheGluAsnSer 277
Qy 865 GAAAGTATACACATATTTTCAAAATCCTAAATCGCGCATAAACAGAGAAAGGTTCT 924
Db 278 SerAspGlyLeuTyrValPheSerAsnGlyGlnCysGlyArgAsnGluIleValPro 297
Qy 925 ACCAATTAATGATATGCTCCCTCAATATTAATTAATTAAGACAGATGGAAGAGATTT 984
Db 298 ThrAsnLeuAspTyrValProGlnHisLeuArgTrpPheAspGluTrpAlaGluAspPhe 317
Qy 985 TGCCGAAAAAGAAATTAATTAATTAATTAAGACGACCTCGTCTGATGCAAAAGAA 1044
Db 318 CysArgLysArgAsnIleThrLeuLysSerAlaLysGluLysCysGlyGluIleLysAsn 337
Qy 1045 CGCTTATATGATGATCAATTAATGACATGATGATGATGATGATGATGATGATGATGAT 1104
Db 338 ThrLysTyrCysSerLeuAsnGlyTyrAspCysThrLysValPheGluLysLysAspSer 357
Qy 1105 TTGCATTTGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1164
Db 358 CysSerSerAspGlyAsnGlyThrAlaCysSerAsnIleCysIleAlaTyrAspAlaTrp 377
Qy 1165 TTAGCGAATCAACAGACATTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1224
Db 378 LeuArgAsnGlnArgAsnGluPheGluLysGlnLysIleLysTyrThrLysGluIleGlu 397
Qy 1225 TCAATTTATGCAACGATTAACAAATTTGCAATTAATTAATTAATTAATTAATTAATTA 1284
Db 398 LysTyrLysSerSerSerAspLysSerAsnSerAsnIleSerAsnLysTyrTyrAsnGlu 417
Qy 1285 TTTTATGAAAAAAGTTAAGAGAACGCAATTAATGCAATTAATGACATTTTAAATTA 1344
Db 418 PheTyrGluAsnPheGlyLysGluTyrGluThrLeuGlnAsnPheLeuLysLeu 437
Qy 1345 AATGAGGAAGATTTGTAAGAGGATTTACAGAGAAAGATTAATTAATTAATTAATTA 1404
Db 438 AsnLysGlyMetTyrCysGlnGluLysIleGlnGluIleValIleAspPhe----- 455
Qy 1405 AGTCGATGATGCAAGGATTTTATCGTTCAATTAATTAATTAATTAATTAATTAATTAAT 1464
Db 456 AsnLysAspGluAspMetValPheHisArgSerGluTyrCysGlnProCysProAspCys 475
Qy 1465 GGGGTCAATGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1521
Db 476 ValValGlnCysLysGlyGlyLysCysThrGluAspLysLysAsnAspLysCysArgSer 495
Qy 1522 -----GTAATTAATGAAGACATTAATTAATTAATTAATTAATTAATTAATTAATTA 1563

Db 496 LysIleIleLysLysIleLeuGlnSerGluLnu-----Pro 507
QY 1564 ACTAATATCATCTGCTTTATAGTGAATGAACAAGCTGATATTACAAATAATGAA 1623
Db 508 ThrGluIleHisValIleuAsnSerAspAspLysGlnIleHisLysLysLeuGln 527
QY 1624 AATTTTGTACAGCTCACTAATATTACAAAGTAAATATCAAAAAATGGAATGCTAT 1683
Db 528 ValPheCysSerSerThrIleAspTyrGlnGluYArgAsnValGlnLysTrpLysCysTyr 547
QY 1684 TATAGAGTGAATAATATATAGTGAATGAAACAAATAATCTGAAATCAATATGAT 1743
Db 548 AsnLysAsnSerAspTyrAsnAsnCysGluMetAsnIleSerSerTyrLysAspSerThr 567
QY 1744 AATCTAAGATATATATCATTTTCAATTTTGTGAATGAGTACATTTATTAAG 1803
Db 568 AspAlaAsnValMetLeuSerValGluCysPheHisSerTrpAlaLysAsnLeuIle 587
QY 1804 GATACATATTAGTGAATGCAACAACTTAATACTGTATTAATATACA--ACCAGCAT 1860
Db 588 AspThrIleLysTrpGlnHisGlnLeuLysAsnCysIleAsnAsnThrAsnValIleThrTyr 607
QY 1861 TGTATTCATGATATGAACAATAATGCTTATGTTTGAAGATGGGTAAACAAGAA 1920
Db 608 CysGluSerLysCysIleLysAsnCysGluCysTyrGlnLysTrpIleLysAspGln 627
QY 1921 GAAGATGCAATGATATAAGAACTGTTCAAAAAAAGAAATATACAGCATGAT 1980
Db 628 HisGluTrpGluLysValIleHisAsnValPheGluAsnAsnAspMetSerTyrIleTyr 647
QY 1981 TATAGTATATATATATATATTTTGAAGTATTTTGAAGTATTTGAAGTATTAAGTTCAC 2040
Db 648 TyrAsnAsnLeuSerAspTyrAlaPheAspSerPheLeuPheGlnValMetPheAlaLeuAsp 667
QY 2041 AAGAGTGA--GCAAAATGGAAGAACTTATGGAATATATAAAGAAAGAAATGAG 2097
Db 668 GlnAspGluLysGlyLysTrpAspGlnPheThrGlnAspLeuLys--LysLysPheGln 686
QY 2098 TTTTCCAAATTTGGAATAATATAGGAGCTATTTAGAGATTTAGAACTCTGTTAGAT 2157
Db 687 ProSerLysThrAsnThrProThrGlyLysSerGlnAspAlaIleGlnPheLeuAsp 706
QY 2158 CACTTAAAGAAAGCTCCACGATATGTAAGCAATATATACAAAGAGCATGTGAACA 2217
Db 707 HisLeuLysAspAlaLeuThrCysArgAspAsnAsnSerAsnGlnLysSerLysAspAl 726
QY 2218 TCCCATTAATGCAACAACACCCGCTGTTAACTCGTGAGGACGCAACCCACTAA 2277
Db 727 SerLysLysValLysThrAsnProCysGlyLysAsnProSerAlaSerAsnLeuVal 746
2278 AATATAAAGAAATATAGCAATATCTTTAAAGAGAGCTAGACAGAGCAAGCAATGCT 2337
Db 747 ArgValLysArgLeuAlaGlnMetGlnArgTyrAlaArgLysGlnLeuGlnLysArg 766
QY 2338 GGTCTTCAT--AAATGAAAGAAAGCAGCAGAGGATATATATAACTGGGGGTAG 2394
Db 767 GlyGlyGlnIleAsnLeuLysLysAlaThrLysGlyThrTyrArgGlnGlyLysPro 786
QY 2395 AGAAGAGCACTCAAGCAAGATTTATAGTAATATATGATTAACATCTC-----AAT 2445
Db 787 AlaAspGlyPheLys--AsnValCysSerIleAsnGlnAsnHisThrAsnValGlnAsn 805
QY 2446 CGTAACTCTGCTTTTCA--AATGACACATGATGCAAA--GGCAGCATGATGCT 2499
Db 806 AsnAsnArgAlaTyrThrTyrGlnGlyProCysThrGlyLysAspGlySerAsnGlyLys 825
QY 2500 ATACAAACAAGATTTGCTGAGCACTAATGGAAGTGCATCCGGAACACATGCGTAA 2559
Db 826 Val-----ArgMetLysIleGlyThrProTrpLys--ProGlyArgGlnIleGlnMet 842
QY 2560 GATCAGCAAGATGATATATGCTCTCTCAAGAGACATATATAGTATCAATTTGGAA 2619
Db 843 SerAlaGlnAspIleTyrMetProProArgArgGlnHisMetCysTrpHisSerAsnLeuGln 862
QY 2620 CATTACAAACGATATACACCACCT--AATGTAATATGTTGATGATTTAGTAA 2676
Db 863 TyrLeuGlnIleThrLysAspGlyProLeuLysGlnIleAsp-----GlyLysLeuValAsn 880
QY 2677 AATTCCTTTTGGGAGATGCTCTATACAGCAAAATATAGCAACAAGATTAATGCA 2736
Db 881 AsnSerPheLeuGlnLysPylAlaLeuSerAlaLysMetAspAlaGlyLysIleLeuGln 900
QY 2737 ATGTATAAAGAAAGATTAACCTAAAGGCCCAAGAAAGAAAGTACTGACCACCAACAG 2796
Db 901 LeuTyrLysGlnAsnAsn-----LysSerAsnLeuThrAspProLysAspAsn 917
QY 2797 ACAACATATCTCGACCTATACGTATACATTTTGCAGATATAGGATATTAATTCAGAGA 2856
Db 918 GluSerAlaCysArgAlaLeuLeuArgTyrSerPheAlaAspLeuGlnLysPylIleArgGly 937
QY 2857 AGAGATCTCTGGGAAAGAAAGCGTACATGTTAAAGCTGCAAGGACATTTGGAACCTGT 2916
Db 938 ArgAspLeuTrpAspLysAsnSerAspAlaLysArgLeuGlnThrAsnLeuLysGlnIle 957
QY 2917 TTTGCTATATATCATATAGTCACTCAAGGCCAAAGAAATGATTAATATATGATGCTCC 2976
Db 958 PheThrLysIleLysGlnGluLeuProGluAspIleLysLysLysTyrAspLysAspGly 977
QY 2977 CCCAAATATTTAAATTTGAGGAAATTTGGGGAAGCTATATAGCCAAATATGAGGA 3036
Db 978 ThrAspHisLysLeuLeuArgGlnLysPyrTrpGlnAlaAsnArgHisGlnValTrpArg 997
QY 3037 GCCATGAATATGATATTAATATATTGAAGTAAATGAGGACCAATCAACACAAAGT 3096
Db 998 AlaMetLysCysAlaIleGlnAsnAspLysAspMetLys----- 1010
QY 3097 AGTTATTCGATATATGATATACCATTTGATGATATATATCCCAAAATTAAGA 3156
Db 1011 -----CysAsnGlyLeuProIleGlnAspTyrIleProGlnArgLeuArg 1025
QY 3157 TGGATGCCGAATGGGCGAATGCTACTGCAAGGTGCAAGAAAGAGATATGAAGTTG 3216
Db 1026 TrpMetThrGlnTrpAlaGlnIleTrpPheCysLysGlnIleSerArgLeuTyrAsnLysLeu 1045
QY 3217 AAGGAGAGTGAAGAGATGATTAAGATTAATGCTCAAGCTGTACGAAGAGAGT 3276
Db 1046 ValAlaAspCysLysSerCysLysGlyLys-----AlaLysSerCysThrGlnLysAsp 1063
QY 3277 GGTACAGGCTGTACGAAGTGCACAGAACTGTATGATGAATATATATATATGATTA 3336
Db 1064 Gly--AspCysThrLysCysLysAlaAlaCysAspAsnTyrAsnLysLysIleLysPro 1082
QY 3337 TGGAAAGCAATGATATATATATATCAAGTAAATACAAAGATTAATGACAGCA 3396
Db 1083 TrpGlnGlnGlnTrpGlnLysIleLysAsnLysTyrAlaGlnLeuTyrLysAlaLeu 1102
QY 3397 ATGCTCTGATTAATATAGTATGAGTTCACAGCTTCCAGTCTCCCAAAATCATATAGACAG 3456
Db 1103 AspSerValAsnGlyLysGlnLysLeuLysLysLysThrAlaSerAspAlaLysGln 1122
QY 3457 AATGTATGAAATTTTCTCGGAATATATACCAACCAAAAT--GGTGGCAAAAGTAAATA 3513
Db 1123 GlnValValHisPheLeuAlaGlnLeuIleArgLysSerGlyGlyLysGlyLys 1142
QY 3514 AGTGATCTAGTGAAGAGCTGATGCTATGCTATACACCAAGTATGAAATGTTGA 3573
Db 1143 AsnValLysThrThrValSerProThrThrProAsnThrLeuTyrSerAlaAla 1162
QY 3574 GCATATCTC--CATGATACAGCAAAATTTGATGATGCTAGTCAACAAATAGTTTGT 3630
Db 1163 GlyTyrIleHisHisGlnLeuGlnArgThrValGlyCysAsnThrGlnLysGlnPheCys 1182
QY 3631 GATCAAAAAGTGAATGCTAAAGATTAACGAAATAATGCTTATAGATTAACACAGAGAC 3690
Db 1183 TyrSerLys-----AsnGlyLysTyrAlaPheLysAspProProLysGly 1197

Db 1711 -----Glycylglutathrinylalanylserinylasp 1721
QY 5633 GCTGCGAATATTTAGACAAACAATTAATAAATTTGTGAATAAAGTGGAGATGT 5892
Db 1722 ThrThrIglutThrValaspGln---GlnLysalaglulaspLysaspGlygly 1740
QY 5693 GAATATAGGTATGAAGAATGTGTCCACACACGCGATTAATGTATAGTCAAAAT 5952
Db 1741 Glu-----ThrProGlnLysGlnThr----- 1747
QY 5953 ATGCCCATCATATAGACATGAACCAAGATTGAAGCAAGTGTATGTCAAGTG 6012
Db 1748 -----GlnProLysValgluValasnProCysasp----- 1757
QY 6013 CCACGAGGTCCACACGCTGTACGAAGGGAACACCGTACCGGATTCATCATATCA 6072
Db 1758 -----LleValLys 1760
QY 6073 AAGCGACGGCATGAAAGAAAGCAAGCAAGCCGCTTCAAAACGCCGAAAGAA 6132
Db 1761 ThrLeuPheThrThrThrIglutThrLeuLysgluAlaCysProThrLysValasnGly 1780
QY 6133 GTGGAAATCTTAACAACAGAAATGCGACACAACAGAACCCGACGAGACACAACA 6192
Db 1781 ArgGlu----- 1782
QY 6193 ACACGAAACGAAACATCAACACACACACAGAAATCTGAGCTGGCAATGTGAAG 6252
Db 1782 ----- 1782
QY 6253 GCCATCTTTCGAAATTAACACAGATACAGGGGTGAATAGAGGTGTATTCAAAACG 6312
Db 1782 ----- 1782
QY 6313 TATGCAATATTCCTTAATGCGGTGTATGTAGGTAACTTAAGAAATGAATGCG 6372
Db 1783 -----LysPheProasnTrpLysCysIle---SerSerGlySeraspLysSerGlySer 1799
QY 6373 ATATGTATGCTCTCTAGGAGAAATAATATATTAATAAATATCAATATTTAAATAT 6432
Db 1800 IleCysIleIleProIleProIleProIleProIleProIleProIleProIleProIle 1819
QY 6433 -----GAAATGAAATAAGCTGACAAATGAT---ATAAAGAGCT 6471
Db 1820 ThrValSerSeraspGlygluThrThrProIleThrHisaspAlaLeuArgLys 1839
QY 6472 TTTATTAATATGACGACATAGAACTCAATTTTGTGTGTTAAATAT---ATAATGA 6528
Db 1840 PheIleGlnThrAlaAlaValgluThrPhePheLeuTrpHisArgTyrLysLysIleLys 1859
QY 6529 AATCTGACGACGAAATGAATGCAAAATGGAACA----- 6564
Db 1860 GluLysgluArgGlnGlnGlnLeuGlnasnGlyThrPheLeuLeuProProAlaGlnLys 1879
QY 6565 -----ATTCCA 6570
Db 1880 ValSerProGluaspasnProGluHisProGlnLysLysLysGlnGlyLysLysIlePro 1899
QY 6571 GATGAATTTAAAGAAATATGATATATATATATATATATATATATATATATATATAT 6630
Db 1900 GlnGlnPheLysArgGlnMetPheThrLeuGlnLysArgLysArgLysValGly 1919
QY 6631 ACT---GATATT-----TCGAATGAATAA 6651
Db 1920 ValLysaspaspValalaglnAlaLeuGlnLysalSerGlyaspasnLysSerGlyaspLys 1939
QY 6652 AAAATTTATATCTGTAACAATATGTAACAACCATCTTCATGAAATATAT--- 6702
Db 1940 AsnIleLysaspIleSerGlnLysIleLysSerValIleGlnLysSerGlnGlnThr 1959
QY 6703 -----AAGAAACGAGATTAACAAAGATGAAGATTAACGTAATATATTT 6750
Db 1960 ProProGlyProLysProGlyGlnThrThrLysProGlnGln-----Trp 1975

QY 6751 TGGAGAAATATTAATAATTTATTTGGAGAGAAATGATATATGATTAATCTATCTC 6810
Db 1976 TrpGlnLysasnGlnLysIleThrPasnAlaMetIleCysAlaLeuThr---Hisasn 1994
QY 6811 ACAGAC-----GAAACGAAAGAA 6831
Db 1995 ThrAspThrArgGlnValaspaspGlnValLysGlyGlnLeuPheGlnLysasn 2014
QY 6832 AAAATTTAGATTAAT---TACAGATACATGACATGACCAATGACGCTTC----- 6882
Db 2015 ThrProLysasnSerGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 2034
QY 6883 -----CTGAGAGGTTGTAAGAAAGCCCAATTTTGA 6918
Db 2035 GlyGlyProIleGlyAsnIleLysLeuGlnGlnPheAlaSerArgProThrPheLeuArg 2054
QY 6919 TGGTTCACAGAAATGGGACAGAAATTTGTATTAAGAGAGAGACAGTGTAAATG 6978
Db 2055 TrpLeuGlnGlnTrpGlyGlnGlnLysPheCysArgLysGlnIleLysLysLeuGlnArgIle 2074
QY 6979 GAGCGGGCTGT---AAGCAATATGAGTGAATGGTAAATGACGCT----- 7023
Db 2075 LysGlnGlnCysHisLysaspGlyLysasnArgasnCysaspaspaspGlyPheGlnCysLys 2094
QY 7024 -----AGACACAGAAATGTGACAG 7044
Db 2095 GluMetCysProCysLysaspGlySerPheGlnThrLeuLysCysProSerCysAlaLys 2114
QY 7045 GCGTGTGAATATCAAAATTTATTAAGAGAGAGAAATGAAATGAAAGCAAAAGA 7104
Db 2115 SerCysLysSerTyrLysLysTrpLysTrpLysSerArgLysLysaspGlnPheThrLysGlnLys 2134
QY 7105 -----GAAAGTTCAAAAGATTAAGATGCGCAAAAG-----TATTAAGATTAAT 7149
Db 2135 GlyAlaTrpGlnLysGlnLysLysaspAlaGlnLysasnAsnAspTyrLysGlnPhe 2154
QY 7150 CTTTCTACTGAAGAGACATAGAGAGACATGCTGCTCATGAAATTTATTAACATGAAA 7209
Db 2155 SerLysThrLeuArgasnTyrAsnaspAlaAla-----PheLeuAsn---Ser 2170
QY 7210 TTAAGAATATATGTCGCAATTAAGATTTCTTGTATGCAAAACCTTTCACAACTA 7269
Db 2171 LeuLys-----AsnGlyProCysSer----- 2177
QY 7270 CCAAAACACACAAACAATCAATCATCCGATGCTAATGATATGCCAGAAATGCTGAT 7329
Db 2178 -----LysasnaspaspaspSerValGlnaspGluIleLys 2189
QY 7330 TATGTTCTCTGAAGA-----TTTAACAGTGTGAGTGTCTGAACTTCAAAAG 7380
Db 2190 PheaspaspGluArgLysThrPheGlyHisGlnThrTyrCysLysProCysSerLys--- 2208
QY 7381 GGATCTATGATTCATACAAAAAAATTTACTGAACCTTAATACCTATGAAATGTGTAGAG 7440
Db 2208 ----- 2208
QY 7441 AAGACACATATATATATCTAAAGAGCAGAAATATATGATATATACCTGAAGAA 7500
Db 2209 ----- 7501
QY 7501 AAATTTTACCTTTGAGTCTACAAAGAAAGAAAGTAATAAATAGTTGACATTAAT 7560
Db 2214 Lys-----Gluasn 2216
QY 7561 AATCCTTGAGTCTTAAGAACTTATGACCTGATTAATATATATAGGAAGAAACCTT 7620
Db 2217 AsnHisCysaspasnSerLysPro-----Asnasp 2226
QY 7621 TGTGAATATGAGAGAAATCTTTAAGGTGATATGAAATGAATGTAACAAAT 7680
Db 2227 Cys----- 2227

QY 7681 TCAAGTCTTATCAGAGAAAAAGAGTATGTCTACCTCCAGACAGACATATGTGC 7740
Db 2227 ----- 2227
QY 7741 TTAAGCAATTTAGATGAATTAATAATGAAGACTTAAGGATGATATATCTCTAAA 7800
Db 2228 ----ATGAsnIleAsnSerIleSerAlaGlu----- 2236
QY 7801 ATGGTTCGTCAACGTCAGCAAAATGAAGAAATAGACATTAATAAAAACTTCAACTCAGAG 7860
Db 2237 -----AspIleGluLysAsnSerAsnSerThr 2245
QY 7861 AAGCGTGCCCAATGAATCCAAATATGTGAT-----ACTATGAATATATACTTCCGCTGAT 7914
Db 2246 GluAspValThrMetSer-----ValSerAspSerAsnThrAsnGluAsnLysPheTyrAsp 2264
QY 7915 CTGGGTGACATAGTATAGAGAACAGATATGTTACGAATGTGGTACTTACCTCCCGTA 7974
Db 2265 LeuAsnAspCysIleLysAla----- 2271
QY 7975 GAAATTAATAATATAGGCTTTTGAATACATATATGMAAATGAGAAATAAAAATMAA 8034
Db 2272 -----GlyIlePheLysGlyLe----- 2277
QY 8035 GGTAGAAATTAATACACGATGTACAAAGCTTTCGTTCTGCTGGTGGATGCTAATAGA 8094
Db 2278 -----Arg 2278
QY 8095 AAGATATTTGGAAGCAATAGCGTCGCAAGACAGAGAGATGCAAAACTTTTAGAAAA 8154
Db 2279 GluAspValThrPheLysCys----- 2284
QY 8155 GGAAGAAATGATGATTTGAACGCATTAATTA-----ATACAGATTAATGCT 8202
Db 2285 GlyLutCysGlyValAspIleCysThrLeuGluLysThrAsnAsnGluLysVal 2304
QY 8203 GGACATAGACGATCCACCTGTTGATGATTAATATACCTCAAGGCTT-----CGA 8253
Db 2305 SerAlaLysGluAsnAspAsnLysAsnGluIleLeuLeuIleArgValLeuPheLysArg 2324
QY 8254 TGCATGACTGAATGCTGATATATTTATGTAAGCACTGATGAGAAATGGAATAATTT 8313
Db 2325 Tripleu-----GluSerPheLysGluAspTyrAsnLysIle 2336
QY 8314 AAAAAATCATGTGATCCTCTAATAAACATGTCACAGATGACAGATGATATGATGAATAAT 8373
Db 2337 AsnAspLysIleSerHisCys-----MetLysAsnAsp----- 2347
QY 8374 AAGCTGCAACAGCTTAACAGACATGTCACAGATATATAAATTTTCTTCTTAATGAAA 8433
Db 2348 LysLysSerProCysIleAsnGlyCysGluAsnLysCysAsnGlyValGluLysTyrPile 2367
QY 8434 TCCTATTCATATACATACAAATTAATTAACAAAGATTTGTAAGAACCAACATATATACA 8493
Db 2368 GluLysLysLysSerGluTyrPglLysValArgLysLysArg 2381
QY 8494 AAAATCTCTATGATGATCATGTTCAAAATTTGTGACAAAAAGTTGAAAAGTTTAAAGT 8553
Db 2382 ----IleAsnGluLys----- 2385
QY 8554 GAATGTTCTGTAGACCTTTTCTGAATATCTTCATGAACAAAGTAAAGTGTGATTAAT 8613
Db 2385 ----- 2385
QY 8614 AAATTTAATGAATATGATGTTCTTCCATATATACAGACATATGCTTTGAGAAACACCA 8673
Db 2386 ----ArgAspLysAsnSerAsn-----GluAlaPheGlu-----Val 2396
QY 8674 AAAAGTATTAAGACAGCTTGACAGTTGATGACATCTTCTTAAGAAATCCATTTGATTAATGT 8733
Db 2397 LysSerPheLysGluLysLeuIleProGluIlePro-----Val 2409
QY 8734 CCTACGATCAAAAAAACA-----GATGATGTAAGAAATTAACAACTTTTACCTTCTGCTG 8790

Db 2410 ValThrAspLysGlyLysHisAspSerLeuThrGluLeuLysLysLeuLysCysAsn 2429
QY 8791 AAGAAATGATATATATATATATATCTTGAATTAATGGAACGATACCTGTTCTTAATAGTCA 8850
Db 2430 CysSerGluLysSerGluAsn----- 2436
QY 8851 GATGATTAACAAGAGTGTATGATTCCTCCAGAAAGACATTTANGTACAGACCTATC 8910
Db 2437 SerAsnGluLysAspValVal-----LeuCys----- 2445
QY 8911 ACTGCATATATATATAGAAAAAGGATAAAGAAATTTTAAAAAACTTCTTACTTCT 8970
Db 2446 -----LeuLeuLysLys----- 2449
QY 8971 GCTTCAGTCAAGACATTTGTTAGTCAAAAAATATTAATGGAAGAGAGTGTGCTT 9030
Db 2449 ----- 2449
QY 9031 GAGGCAATGAATATAGTATATGAGATTAATCCGATTAATTAAGAGAACTGATATGATG 9090
Db 2449 ----- 2449
QY 9091 GACACTTCATATCTGAAAAAATTAATAATATTTGAACATCAATGAAGACCGAA 9150
Db 2450 ----LeuGluAspLysAlaLysAsn----- 2456
QY 9151 AATCGTAAACATGATGGGAAAAATTAATAGACGTGATATGCGACGCTATGTTATGGA 9210
Db 2456 ----- 2456
QY 9211 TATTAATTTGCTACTTCAAAAGTAACATTAAGACAGATGGTCTCAATACCAAGAT 9270
Db 2457 -----Cys-----LysAsp 2459
QY 9271 GAAGAACTAATCAGTTCTTCTGTTGTTAATGAATGGCGAAAGCAAGCATGAAGAA 9330
Db 2460 GluAlaSerGlyGlu----- 2464
QY 9331 AAGAAACATGTAAGTATTCATTAATAAACAAATGCTCTGTCACAAAGAAATATTTT 9390
Db 2465 -----ProCysProGlnThrSerGluAsnPro 2474
QY 9391 GAAGCTCAGAAATTAATTAAGACAACTGATGTCCAGATGATTAATTAATATATAGC 9450
Db 2475 AspAspLysPheLysLeuLeuGlu----- 2483
QY 9451 TTGAATATATGATTAATAATTAACAAAGTAAGTAATTAATTAATTAATTAATTAATTA 9510
Db 2484 -----GluAsnProValGluAlaProAsnIle----- 2492
QY 9511 GATCAATCTTCAGGTATATATAGACAAATTAACCAATCTGAAAGAAATGTTCAAGTCATATATA 9570
Db 2493 -----CysProLysValGluGluProGluProValVal 2503
QY 9571 AAATCAAAAGATTTCTCATGCGCTTGGAGTTAAATGATTAATAATGAATAGTTACAGCA 9630
Db 2504 ----GluLysGluLysCys-----AspLeuAlaGluAlaProSerLys 2516
QY 9631 ACAAAAAATATATATATATATATATCAAGAGTACTTAATAAAATATATATATATATATATAT 9690
Db 2517 GluSerSerThrGluGluAsnSerGlyGlu----- 2526
QY 9691 TATTTGTTGAAGATGAACACAAACAAACAAATCATGATAGATGAATATTAAGAGAA 9750
Db 2527 -----GlySerAsnSerGluLysAsnProArgSerLysProGlu 2539
QY 9751 GAGCAAAACAGTGTCTCTAAAGACATATTTCTTTACACCCCATGTAGATCTTTCTAT 9810
Db 2540 GluGluLysProProProLysSerGluThrAspThrPro----- 2553
QY 9811 CAAGCATTATATCTCAACACATCGAGTAGCAAAATATGAT-----CCTAAAAAT----- 9861

Db 2554 ---ProFolalaproProthrillegInProSerGlnlaaspInProthraenSerile 2572
 Qy 9862 ---GATATATTGAAGTAGTACTCTGTTATTGTATGCGGCTAGGTGGATACGC 9918
 Db 2573 SeraspIleuSerSerThrIleProPhcGlyIleAlaIleAleuThrSerIleVal 2592
 Qy 9919 CTTCATTTCATGAAGAAAATTCAAATCGCTGTGACTTGTGGATATGCAATATC 9978
 Db 2593 PheLeuPheLeuLysLysLysThrLysSerSerValaspLeuLeuValLeuAsnIle 2612
 Qy 9979 CCCAAGAGAGATGAGTAAGCTAGCTGAGTTCGAAATAGTGACATACCATATAGA 10038
 Db 2613 ProLysGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2632
 Qy 10039 AGTGGTCATATTAAGCAAAACATATATATATATATATATATATATATATATATAT 10098
 Db 2633 SeraspArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2650
 Qy 10099 GATAAATATATGTGGACCTATCTCTGATATATATATATATATATATATATATAT 10158
 Db 2651 GlyHisLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2668
 Qy 10159 GAATGGAT 10218
 Db 2669 GluLeuaspIleAsnaspIleLysLysLysLysLysLysLysLysLysLysLysLys 2688
 Qy 10219 GATAGTACTGGAACCATCAAAAGGCAATACCAAGTAT---GATACACCAACT--- 10269
 Db 2689 ValValLeuGluLysProSerLysAsnThrThrAlaSerGlyLysAsnThrProSer 2708
 Qy 10270 ---AATGATACACCACTAGCAATAGATATATATATATATATATATATATATAT 10311
 Db 2709 ThrGlnAsnaspIleGlnAsnaspIleProSerSerLysLysLysLysLysLysLys 2728
 Qy 10312 AATGACATGAACATGATTTGTATCTCAATATATATATATATATATATATATATAT 10365
 Db 2729 AsnThrLysaspIleLysLysLysLysLysLysLysLysLysLysLysLysLys 2748
 Qy 10366 ---AATATATGAAGAGCAATATATATATATATATATATATATATATATATAT 10422
 Db 2749 ProAsnaspIleLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2768
 Qy 10423 GATAAATCTGGAAGAAAACCTTTATATATATATATATATATATATATATATATAT 10482
 Db 2769 HisAsnValGluGluLysProPheIleMetSerIleHisaspLysAsnLeuThrGly 2788
 Qy 10483 AAGAATATAGTAT 10542
 Db 2789 GluGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2807
 Qy 10543 GCTAGAATGATCTTATAGAGTATATATATATATATATATATATATATATATAT 10601
 Db 2808 ArgaspAsnAsnValLysSerGlyLysLysLysLysLysLysLysLysLysLysLys 2825
 Qy 10602 CCTATGAT 2828
 Db 2826 AspTyraspIleLysaspGluValLeu 2834
 RESULT 3
 T28431
 Variant surface protein 1 homolog A4VAR - malaria parasite (Plasmodium falciparum) (frag
 C:Species: Plasmodium falciparum
 C>Date: 15-Oct-1999 #sequence-revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T28431
 Mol. Biochem. Parasitol. 97, 133-148, 1998
 A:Title: Analysis of adhesive domains from the A4VAR Plasmodium falciparum erythrocyte
 A:Reference number: Z20486; M01D:99094502; PMID:9879893
 A:Accession: T28431
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3026 <SMT>
 A:Cross-references: EMBL:LA2244; NID:93540144; PID:93540145; PIDN:AAD03351.1

C:Genetics:
 A:Gene: var
 A:Introns: 2906/3
 Alignment Scores:
 Pred. No.: 8,256-190
 Score: 4010.00
 Percent Similarity: 42.808
 Best Local Similarity: 29.768
 Query Match: 20.61%
 Gaps: 127
 US-10-087-013-1 (1-10628) x T28431 (1-3026)
 Qy 88 GAAAGTCACAAAGTCGCAAGAAATGTTTGGACGTTATGCCAAAATATA---AGACAT 144
 Db 13 GluaspaspLysaspAlaLysHisValLeuaspSerIleGlyGluValLysLysGlu 32
 Qy 145 CCATCAAAATATATGCAAAAGAACATGTGATTCGTGAAAGGAGTTTGCAGAAAGCAAA 204
 Db 33 ValLysLysValAlaAlaAsnLysSerGlnLeuLysGlyThrLeuSerAsnAlaIle 52
 Qy 205 TTTCGTGCTGCTCTTCTACCGCACTAAATAGCATATATATATATATATATATATATAT 264
 Db 53 PheGluaspGluProLysGlyGlnGlnThrGluAsnasp-----ProCys 67
 Qy 265 AATTTAGATCATTAAGCAACATACATATATATATATATATATATATATATATATATATAT 324
 Db 68 LysLeuLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 86
 Qy 335 TGCCATGCTAGAGAAACAAACCGATTGATGATGATGATGATGATGATGATGATGATGAT 381
 Db 87 CysargLysGlyThrGluLysArgPheSeraspValGlyGlyGlyGlyGlyGlyGlyGly 106
 Qy 382 AAAATACGAT 441
 Db 107 LysLysLysasp-----SerLysAsnaspGlyLysLysLysLysLysLysLysLys 124
 Qy 442 CATATGCTGATTAATAAAGCTTGAAGCTCTTAATGATATATATATATATATATATATATAT 501
 Db 125 HisLeuGlyValAlaGlnLeuGluAsnIleSerAlaLeuaspLysLysAsnaspThr 144
 Qy 502 TTATTTGGAAATGATACAT 561
 Db 145 LeuLeuLysaspValCysLeuAlaAlaLeuHisGluGlnSerIleThrGlnAspTyr 164
 Qy 562 CCACATGAAA-----GGAACCTCAGACGCTTGACTGCTCTT 597
 Db 165 ProLysLysGlnAlaGlnLysLysSerPheSerProSerGlnLysThrMetLeu 184
 Qy 598 GCACGAAGTTTGCAGATATAGTATATATATATATATATATATATATATATATATATATAT 657
 Db 185 AlaArgSerPheAlaAspIleGlyAspIleIleArgGlyLysaspLeuLysIleGlyAsn 204
 Qy 658 GTCATGACAAA-----GTAGAAGCGGTCTCCGAGAGTTTTCAGAAATATATATATAT 711
 Db 205 LysLysGluLysLeuaspLeuGluLysAsnLeuLysLysLysLysLysLysLysLysLys 724
 Qy 712 GCAATGCAAGAT---GAAGTAAAAAATGATTAACATCCGATGATGATGATGATGATGAT 768
 Db 225 LysLeuThrAspProArgAlaLysAspHisLysLysLysLysLysLysLysLysLysLys 244
 Qy 769 AAATTAAGACAGACAGCTGCAATATATATATATATATATATATATATATATATATATAT 828
 Db 245 GlnLeuArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 264
 Qy 829 GATGATCAT 888
 Db 265 HisAlaGlyGluSeraspLysLysLysLysLysLysLysLysLysLysLysLysLys 284
 Qy 889 AATCCTAATGCGCGCATAAACAAGAAAG-----GTCCCTACC 927
 Db 285 GlnGlyArgCysArgCysAsnaspLysLysProGlySerAsnThrAspProThr 304

QY 928 AATTAGATATATGCTCCATATTTACGTTGGTGCACAAATGGGAGAAAGTTTGC 987
DB 305 TATPheAspTyrValProGlnTyrLeuArgTyrPheGlnLutrpSerGluAspPheCys 324
QY 988 CGAAAAAGAAATTTAAATTTGAAAAAGGTCAAGACTCCGTCGT-----AAT 1035
DB 325 ThrLysArgLysHisLysLeuGluAsnAlaIleGlnLysCysArgLysGlnAspGlySer 344
QY 1036 GACAAAGAACGCTTATATTTAGTATCATATGACATGATGTGTACGACAACTATTTGGAAA 1095
DB 345 GLLysGlnLutry---TyrCysAspPheAsnGlyPheAspCysLysGlyThrAlaSerGly 363
QY 1096 AAGGATATTTGCAATTTGATTAATTAAGTACTGACTGTTGCACTAAATGCAAAAGTTT 1155
DB 364 LysHisLysTyrLeuTyrPheAspAsnAlaCysAlaGlyCysPheSerCysSerAspPhe 383
QY 1156 GAAGTTGGTTAGGATGACAAAGAACGATTTAAAGAAAGAAAGAAATATGAAAAA 1215
DB 384 ArgLysTrpIleAlaLysGlnLysAspGluPheGlnLutryGlnLysAsnLysTyrThrAsp 403
QY 1216 GAATATCAATCATATTTATTCGACGATACAAATTTGTC-----AATATATTT 1263
DB 404 GlnIleLys-----LysAsnAspAspThrThrIleThrGlnLutryGlyThrIle 420
QY 1264 AATATGCAATATTTATTAACAATTTTATGAAAACTTAAGAAACGCAATATGCACTAAT 1323
DB 421 AsnAsnMetLysTyrArgLysAspPheTyrLysHisLeuGlnLutry---LysTyrLysThrVal 439
QY 1324 GACACTTTTAAATTTACTTAATTAAGAAAGATTTGTAAGAA-----GGA 1371
DB 440 AsnAlaPheLeuAsnLutryGlnLysGlnLutryGlnLutryGlnLutryGlnLutryGln 459
QY 1372 TTACACGAGCAAAAGATATTTACTTTTACTTAACAGCTGATGCAAAAGGATATTTTAT 1431
DB 460 GlnGlyLysLysLysTyrIleAspPheAsnAspAsnIleGlu-----ThrPheSer 476
QY 1432 CGTTCAAGATATTTGCCAAGTGTGTCCGACTGCGGGGTCAAAATGTGATGATTAATATAC 1491
DB 477 HisThrLutryArgGlnLutryProCysProTyrCysGlyLutryGlu----- 490
QY 1492 ACACACAAATCAGATATGATCGACGCTGAATATAT----- 1530
DB 491 ---GlnGlnLysAspGlyLysTyrPheArgIleAsnAspHisSerAlaCysLysGlnLutry 509
QY 1531 GAAGACTATTAACCTCCATGCGGGGTGAGACCTATATATCACTGCTCTTATATGCT 1590
DB 510 GlnLeuTyrThrProLysGlnLutryGlnLutryThrLysIleAsnValLeuThrSerGly 529
QY 1591 AATGACAAAGCTGATATTTACACAAAATTTTGAATACTTGTACAGCTCAACTAATAC 1650
DB 530 GlnGlnLysGlnLutryGlnLutryGlnLutryGlnLutryGlnLutryGlnLutryGln 549
QY 1651 AAGATATTAATATAT-----CAAAAATGGGAA 1677
DB 550 GlnGlyLysSerAspAspCysGlyLysAsnSerAspSerSerLeuCysGlnLutryGln 569
QY 1678 TGTATATTAAGATGA----- 1695
DB 570 CysTyrGlnProAspGlnLeuGlnLutryGlnLutryGlnLutryGlnLutryGlnLutry 589
QY 1696 ---AATATTAATGATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 1737
DB 590 GlnAlaGlyLysLeuCysIlePheGlnLutryGlnLutryGlnLutryGlnLutryGlnLutry 607
QY 1738 AATGATATCTAGAT 1797
DB 608 ---LysGlnLutryThrPheAsnAsnPhePheAsnPheTyrPheAlaHisVal 623
QY 1798 TTAAGGATATCTATTAAGGATGACAAACTTAATACTGTATATATAT-----AATACAAAC 1854
DB 624 LeuLysAspSerIleAspTyrArgThrGlnLutryGlnLutryGlnLutryGlnLutryGlnLutry 643

QY 1855 ACCGATTTGTTATGATTAATGTAATTAACAGAAATGCTATGTTTACAGATGCGTTAAACA 1914
DB 644 LysLysCysGlnLutryGlnLutrySerAsnCysGlnLutrySerLysTyrIleGlnLutry 663
QY 1915 AAGAAAGAAAGATGATGATATATTAAGAAAGCTGTTCAAAAAGAAAGAAAT-----ATACAG 1971
DB 664 LysGlnLutryLutryPheLysValLysAspGlnPheAsnLysGlnLutryAspPheLeuGln 683
QY 1972 CAATCGTATTTATGATATATTAATTAATCTTTTGAAGGTTATTTTAAAGTTATGAT 2031
DB 684 TrpLysHisTyrLeuValLeuGlnLutryIleLeuGlnLutryGlnLutryGlnLutryGln 703
QY 2032 AACTTGCAAAAGATGACAAAGAAATGAAAGAACTTATGAAATATTAATAAGAAAAA 2091
DB 704 LysAlaTyrGlnLutryLutrySerIleGlnLutryGlnLutryGlnLutryGlnLutryGln 723
QY 2092 AATGATTTTCAATTTGAAATTAATAGGACATTTATGAAATGCAATGAAAGAACTGTC 2151
DB 724 GlnAsnLysAsnArgThrLysAspAsp-----GlnAspAlaLeuAspValLeu 739
QY 2152 TTAGATCACTTAAAGAAAGAACTGCCAGATATGT-----AAAGACAAATAT 2196
DB 740 PheAspHisGlnLutryGlnLutryGlnLutryGlnLutryGlnLutryGlnLutryGlnLutry 759
QY 2197 ACAAGACAGACATGTAACATCCATATGCAACAAACCAACCGCTGTGTTAAACCTCGT 2256
DB 760 AspAspLutryGlnLutryGlnLutryGlnLutryGlnLutryGlnLutryGlnLutryGln 775
QY 2257 GGAGGACGACACCACTAAATTAATTAAGAAATATGCAACAAACTTAAAGAGATGCA 2316
DB 776 SerGlyThrArgHisArgAlaMetValLysAsnValAlaAlaAspMetTyrGlnAlaAla 795
QY 2317 TACGAGAGACGCAAAATCT-----GCTCTCATTAATTTGAAGAAAGGACACAGAA 2370
DB 796 ArgGlnGlnLutryGlnLutryGlnLutryGlnLutryGlnLutryGlnLutryGlnLutry 815
QY 2371 GGTATATATTAAGCTGGGGGTAGAGAAAGACTTCAAGACATTTATGATTAATATG 2430
DB 816 GlnHisTyrAsnGlnLutryGlnLutryGlnLutryGlnLutryGlnLutryGlnLutryGln 834
QY 2431 ATAAACATTTCTAATCTGTAATCTTTGTTTCAAAATGACATGATGCAAGAGACACA 2490
DB 835 AsnGlnLutrySerAsnAla-----IleGlyAspSerLysAspProCysAsnGlnLutryGlnLutry 853
QY 2491 GGTGATGCTATACAAACAGATTTGCTGTAAGAACTGAATGGAATGATCCGGAACAC 2550
DB 854 GlnLutry-----PheLysIleGlyThrProTyrPheAsnIleValLysLys 867
QY 2551 ATGCGTAAGATCAGAAAGATTTATATGCTTCTTCAAAATGACATGATGATATATGATCC 2610
DB 868 LysThrThrSerTyrLysAspValPheLeuProTyrArgTrpGlnLutryGlnLutryGlnLutry 887
QY 2611 AATTTGCAATTTTACAAAGGATGATCACCCTTAATGATGATATATGTTGATGATTTA 2670
DB 888 AsnLeuLutryGlnLutryGlnLutryGlnLutryGlnLutryGlnLutryGlnLutryGlnLutry 902
QY 2671 GTTAATATCTCTTTTGGGGGATGTTCTCTATCAGCAAAATATGAGCAAAACAGATA 2730
DB 903 ValAsnAsnLysPheLeuValGlnValLeuLeuSerAlaAsnLysGlnAlaGlnLutryPhe 922
QY 2731 ATGCAATGTAATAAGAA-----AAGATTAAGTAAAGGCCCCCAAGAAAGTAACT 2781
DB 923 LysGlnLutryGlnLutryGlnLutryGlnLutryGlnLutryGlnLutryGlnLutryGlnLutry 938
QY 2782 GACCAAAACACGACAGACTATCTGCGATATGATGATGATGATGATGATGATGATGATGAT 2841
DB 939 ---CysArgAlaLeuLysSerSerPheAlaAspLeuGly 950
QY 2842 GATTAATTTGAGAGACATCTCTGGAAGAAAGAGCGTACATGTAAGCTGCAAGCA 2901
DB 951 AspIleLutryGlnLutryAspLeuTyrPheAspLysAspSerGlnLutryGlnLutryGlnLutry 970
QY 2902 CATTTGAAACTGTTTGTGTAATATATATGATCACTCAAGCAAGCAAGAAATGATATA 2961

Db 971 AsnLeuValThrIlePheGlyLysIleValGlnArgLysGlyIleAspThrSerLys 990
QY 2962 TATAATGATGATGCCCCCAATATTTAAATTTGAGGAAATTTGGAGAAATTTGAGAAATGAGA 3021
Db 991 TgtThrAsnThrAspGlyLysHisAsnGlnLeuArgGlnSprTrpGlnAlaAsnArg 1010
QY 3022 GCCAAATATGGGAGAACCCATGAAATGTGATATAAATATTTTGAAGATAAATCGGCACAC 3081
Db 1011 ArgGlnValTrpLysAlaMetLysCysAlaLeuLys 1022
QY 3082 CAATCAACACAAAGTACTTATTTGGGATATAGATATACCATTTGATGATATATC 3141
Db 1023 -----GlyGlnLysIleAsnCysGly-----AlaThrProLysAspArgLys 1037
QY 3142 CCACAAAATTAAGATGATGATGACCCGAAATGGCAGAAATGTGACTCAAGGTGCAAAAGAA 3201
Db 1038 ProGlnArgLeuArgTrpMetThrGlnTrpAlaGlnTrpPheCysLysGlnInserArg 1057
QY 3202 GAGTATGATTAAGTTGAAGGAAAGTGAAGACTGTAAAGATTAAGATTAAGTCAAGC 3261
Db 1058 LeuTrpAspLysPheLysMetGlyLysCysGlyLysCysMetAsn-----Gly 1072
QY 3262 TGTACAAAGAGAGGTGATCAGTTGATGACAGTGCACAGTGCACCTGTATGATTAAT 3321
Db 1073 IleCysAsnLysValLysAspAspCysAlaLysCysThrGlnAlaCysLysGlnTrpLys 1092
QY 3322 GATATTAATAGATTAATGAAAGAAACAAATGATATATATAGATTAATATC-----AAA 3375
Db 1093 ThrLysIleGlnProTrpLysAspGlnTrpGlnLysLeuGlnLeuLysIleValLeuSer 1112
QY 3376 GAATTACATGAACACACAAATGTCTGTATGATTAATGATTAATGATTAATGATTAAT 3435
Db 1113 TyrLeuHis----- 1115
QY 3436 GCCAAATATCATATAGACAG-----AATGTT 3462
Db 1116 AlaLysAsnAspSerArgTrpMetAlaPheGlyLysThrAspProAspTrpGlnGlnVal 1135
QY 3463 ATTGAATTTTGTCCGATATATACCAACA-----AATGTCGCAAAAGTAAATGAAGTGT 3519
Db 1136 ValHisPhePheLysGlnLeuGlnAlaIleLysSerSerThrSerLysArgProLys 1155
QY 3520 ACTAGTATGAAGTGTCTCATTTGATGATGATGATGATGATGATGATGATGATGATGAT 3579
Db 1156 ArgSerThrAspAlaIleThrThrAspProThrTrpProLysSerThrAlaIleGlyLys 1175
QY 3580 CTCATGATATACGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3639
Db 1176 IleHisGlnGlnIleGlyAsnAlaGlnLysGlnIleGlnLysHisPheCysAspAsn 1195
QY 3640 AGTATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3699
Db 1196 LysAsp-----LysTyrValPheArgGlnLysProLysAspHisAspGln 1210
QY 3700 GCGTGTGTTGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3759
Db 1211 AlaCysAsnCysThrGlnAsnValGln-----LysProLys 1222
QY 3760 GCGGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3810
Db 1223 LysGlnGlnLysIleAspThrLysLeuAspValCysAlaIleValLysThrAlaLeuThr 1242
QY 3811 AACGATGAGAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3870
Db 1243 GlnAspAsnLeuThrLysAla-----CysGlnGlnLys-----TyrGlnHisPro 1257
QY 3871 GAT-----TGGCAATGC-----GGAAATATAAATTTAGTGAAGACCCCT 3909
Db 1258 GlnArgHisTrpGlnTrpLysCysIleSerGlnThrThrLysSerSerAspSerGly 1277
QY 3910 CGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3963
Db 1278 SerIleCysValProProArgArgLysLeuTyrValThrProLeuThrLysTrpAla 1297
QY 3963 ----- 3963
Db 1298 GlnGlnAlaThrGlnGlnLysProThrSerProGlnAlaGlyGlnValThrLeuProThr 1317
QY 3964 -----GATTAATGAAATTAATAAATTAATAAATTAATAAATTAATAAATTAATAAAGCT 4011
Db 1318 ProAlaThrAlaSerSerGlnAlaProLysGlyAspSerLeu-----LeuLeuThrAla 1335
QY 4012 TTCATCAAAATTCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4071
Db 1336 PheIleGlnSerAlaAlaValGlnThrPhePheLeuThrPheLysLys-----MetAsp 1354
QY 4072 GGTGAAGAAATGAAGTGTGATTAAGATTAATAAAGAAAGCAGCAGCAGCAGCAGCAGCAG 4131
Db 1355 AsnAsnGlyLysPheAlaLysPheLysLeuLysLysGlnLysIleProGlnLysPheLys 1374
QY 4132 AGATCCATGTTCTACACATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 4182
Db 1375 ArgGlnMetPheTyrThrLeuAlaAspTrpArgAspIleCysIleGlyValLysGlnAsp 1394
QY 4183 -----GATATATCAAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 4224
Db 1395 ValIleLysAlaLeuGlnValaSerSerAspAsnLysSerGlyAsnAsnIleLysGlnLys 1414
QY 4225 -----CAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4272
Db 1415 SerAspLysIleGlnAlaIleLeuLysGlnInserGlySerLysLeuProGlnLysLeuPro 1434
QY 4273 -----AAAAACGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4314
Db 1435 ValThrProAsnAsnValLysAsnProLysThrTrpArgGlnAsnAlaHisIle 1454
QY 4315 TGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4359
Db 1455 TrpHisGlnMetValCysAlaLeuThrTyrLysGlnAspThrGlnGlyAlaLysGlnLys 1474
QY 4360 GATGATTTTACCGAAAC----- 4377
Db 1475 ThrSerIleThrGlnAspProThrAlaIleGlyLysLeuThrAspAsnAspGlyLysLys 1494
QY 4378 -----TACGTTACAAACGTCAAATTTAGT----- 4404
Db 1495 ProLysGlnAspLysTyrAspTrpLysIleValLysIleSerSerValProSerSerAsn 1514
QY 4405 -----GACAAACACACACTTTGGAGAAATTTGCCAAACACACCCACTTTTTCGATG 4458
Db 1515 ValProSerGlyAspThrLysLeuGlnLysPheSerArgArgProThrPhePheArgTrp 1534
QY 4459 CTAACGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4518
Db 1535 LeuGlnGlnTrpGlnGlnLysPheCysArgGlyArgLysAspLysGlnLysIleGln 1554
QY 4519 GAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4545
Db 1555 LysGlnLysTyrGlnLeuAsnTyrArgGlyHisArgGlyLysSerGlyAspGlnHis 1574
QY 4546 AAGTGTGAT-----ACAGAA 4560
Db 1575 IleCysGlnLysThrAspThrSerArgAsnAsnThrPheIleAspLeuHisCysProArg 1594
QY 4561 TGTAAATGAAATGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4617
Db 1595 CysLeuLysGlnLysIleLysTyrLysArgTrpIleGlnLysLysGlnLysGlnLys 1614
QY 4618 CCACAAATATAATTTCAAGATGAAGCGCAGCAAAAGAAATTTGATGATGATGATGATGATGAT 4677
Db 1615 AsnGlnLysAsnAsnTyrGlnLysGlnPheAsnAsn-----AspLeuLysGlnLys 1631
QY 4678 GGTGTAAATGTTACAGCAGTACTGAAACGATGCAACAGATTAATCTTGAACAGCAATTT 4737
Db 1632 Gly-----TyrSerSerPheAsn-----AsnPhe 1639

```
QY 4738 ACTGCTAGTGTGTGAT---AACCTGGAATGCTCTGTGTGACAAAGAAAT---ATA 4791
    |||||
Db 1640 LeuAlaSerLeuAsnHisCysLysHisGluHisAsnArgAspLysAsnLysIle 1659
    |||||
QY 4792 CAATGTGTAGAAAACAGGCTTACTGATGATGCGACAAACATGT---GGGTGCACAAAA 4848
    |||||
Db 1660 GluPheAsnAsnHisThrLysThrPheGlyProSerGluTyrCysLysAlaCysProVal 1679
    |||||
QY 4849 TTTATGTGAAAATGACGACAAATATACATTTTCGAGTAAGATTAAGTCAAGATTA 4908
    |||||
Db 1680 TyrGluValLysCysAsnLys-----LysAsnGluGluCysGluThrIle 1694
    |||||
QY 4909 GTTAAGAGGACAAACACAGTGTCTTAAGTGCACAAAACAGGTCTCTAATACTACAT 4968
    |||||
Db 1695 HisLysThrAspLeuAsnGly-----GluAsnAspAsnAsnTyrThr 1708
    |||||
QY 4969 AACTGTGAAAAGATGAGTAAGATGTGCTTTTCTCTCTGCTGACACTAATATGTTT 5028
    |||||
Db 1709 AspIleLysValLeuValIleAsp----- 1716
    |||||
QY 5029 CATGCAATTGGATGCAATTATACAGATCCAGAAATTAAAGATGAAAATGGCTTCCGAAAA 5088
    |||||
Db 1717 -----ArgLysGluGluSerAsnAspGluGluLeuLysAsnValCysAsnAsnThrSer 1734
    |||||
QY 5089 AGATTGATGAAAGTGGCGGACGACGAAAGGTACAAATTTGGGTCAATCTCAAAAGAAAA 5148
    |||||
Db 1735 LeuPheLysAspSerSerVal-----GlnTyrTyrLysCysGln 1747
    |||||
QY 5149 AAAGAAAAAGAG-----AAATAAAAAAGTCGGATGGCCACAA 5187
    |||||
Db 1748 LysLysAsnGluValAspGlnCysIleIleAspAsnPheLeuAspIleAspLys 1767
    |||||
QY 5188 TATCTTATAGAGTCCCGCTGTAGTGTATGATGATGATGATGATGATGATGATGATGAT 5247
    |||||
Db 1768 Tyr-----MetGluPheAsnValPhePheGlnArgTyr 1778
    |||||
QY 5248 ATATATCTAGCTATGTAATTTTGAAGATGAAAACAAAGACCGAGAAATTTGAAG 5307
    |||||
Db 1779 LeuArgTyrPheValHisAspTyrAsnIleLeuLysAspLysIleLysProCysIleLys 1798
    |||||
QY 5308 AAATATTTTAACAAAATGACATCAGTGGCAAGGAGATGATGATGATGATGATGATGAT 5367
    |||||
Db 1799 -----ThrLysAspGluLysSerAsnLysCysIleAsnGlyCys 1811
    |||||
QY 5368 CCGGATGACTCCGCAAAATTTTCTGACAGCAAAATGAGCAATCTGTGTGAACGCA 5427
    |||||
Db 1812 LysGluLysLeuGluCysValLysLysTyrLeuLysGlnLysGlnAspGluTyrLysAsn 1831
    |||||
QY 5428 ATGATATGCGGGTACAAACCTGTAGGATGATGAAATAGTGA----- 5472
    |||||
Db 1832 IleLysAspHisTyrGluLysAsnLysSerLeuTyrGluTyrGluIleProHisTyrVal 1851
    |||||
QY 5473 -----AATAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 5517
    |||||
Db 1852 LysSerTyrPheValGluGlnLeuTyrPheAspLysAspTyrLysLysAlaGlnAspVal 1871
    |||||
QY 5518 CCTTCAGATGATGATATATCTATGGGAAAAAATCGCGATGAGTACTGCTATCAGTTT 5577
    |||||
Db 1872 IleGluAsp----- 1874
    |||||
QY 5578 CTTCGATGCTTCCCAATGGGTGAAGATTTTTCGAAACATTAAGAAAAGAAATTTGAG 5637
    |||||
Db 1875 -----GluAsnGluArgLys 1879
    |||||
QY 5638 AAATGTGATGGGCGTGTATGATTAATCTTGTGTGATATGAAAGTAAAGAAAAGAAA 5697
    |||||
Db 1880 LysIleTyrPheLysThrAspGluValGluCysThrAsnGluGluTyrLysGluAsnLys 1899
    |||||
QY 5698 TGTACAGATCGTGTACACAAATTAATAAATTTATAGTAGTGAACACACATATGAA 5757
    |||||
Db 1900 -----AspPheIleThrAsnLeuIleLysLysLeuGlnIleLysIleGluSerCysGln 1917
    |||||

QY 5758 AAACAAATCAAAAATATGTGTGAGATTAAGACAAAATATATTTCCGACATCTGTGGCA 5817
    |||||
Db 1918 ThrGlnHisAsnProAsnGly-----LysThrLysProCysAspGluIleProPheHis 1935
    |||||
QY 5818 AAGATGCAACAGACGCGTCCGAATATTTGACAAACAAATTAATAAATTTGTGAAT 5877
    |||||
Db 1936 SerAsp-----GluThrLeuAspGluGlnThrAspThrThrAspAsp 1950
    |||||
QY 5878 AAAGTGAAGATTTGTGAATTAAG-----TGTATGAAGAATGTGTCCACA 5922
    |||||
Db 1951 AspAsnSerAspLysIleTyrAspThrLysProProPheCysProLysAspValGluAsp 1970
    |||||
QY 5923 CACCGATTAACGTAGTGAATAGTCAAAATATGCCCGCATCATTAAGCATGAACAAAA 5982
    |||||
Db 1971 ThrLysGluThr-----GluLysProLys 1978
    |||||
QY 5983 GAAGTTGAAGAAAGTGAATATGTAAGTCCAGACGAGTCCACAGCTGATGACGAAGGAA 6042
    |||||
Db 1979 -----ValLeuProGluProPro----- 1984
    |||||
QY 6043 ACACCGTACACGCGGTATCATGATATCAAAAACGACGCAATGAAAAAGAAAGGAAA 6102
    |||||
Db 1984 ----- 1984
    |||||
QY 6103 ACAGCCCGCTTACAAAACACCGGAAAAAGTGAATACTAACACGAATGCGAGCA 6162
    |||||
Db 1984 ----- 1984
    |||||
QY 6163 CAACAACGAAACCGACGACGACACAAACAAACGAAAGCAATCAACGACAAACA 6222
    |||||
Db 1984 ----- 1984
    |||||
QY 6223 ACAGATCTGACGTGGGACAAATGTAAGGCCATTTTGAATAAACCGATATGACAG 6282
    |||||
Db 1985 -----AspLacCysGluIleValGluGluIleLeu-----AsnGluAsnAspGluTyr 2000
    |||||
QY 6283 GGTGGAATGAGGCTGTATCCAAAACGTA-----GACAAATCTCTAATAGG 6333
    |||||
Db 2001 LysLysIleGluLysCysAsnThrLysTyrTyrProThrLysAsnAspTyrProGluTyr 2020
    |||||
QY 6334 GGTGTATGTAGTAAAGTAAAGAAATGAAATGGAATGCAATGATGATGATGATGATGATGAT 6393
    |||||
Db 2021 AsnCys-----ThrAspLysValIleAsnArgGlnGluGluLysSerMetProProAspArg 2039
    |||||
QY 6394 AAAAATATGATATTAATAATATATATATATATATATATATATATATATATATATATATAT 6453
    |||||
Db 2040 GlnLysLeuGlyLysIleHisAsnLeuGlnHisLeu-----SerGluLysAlaThrGln 2056
    |||||
QY 6454 AATGATATTAAGAAGCTTTATTAATTAATGTCAGCAATGAAACTCAATTTTGTGGTTA 6513
    |||||
Db 2057 ThrGluLeuArgLysAlaPheIleGluCysAlaIleIleGluThrPheTyrPheTyrAsp 2076
    |||||
QY 6514 AAATAT----- 6531
    |||||
Db 2077 LysTyrLysGluAspLysLysAspGluLysLysThrGluGluGlyLysIleSerAspAsp 2096
    |||||
QY 6532 CCT---GCACGAAATAATGATTTGCAAAATGAAACATTCAGATGAAATTTAAAGATA 6588
    |||||
Db 2097 ProAspAspProGlnLysLysLeuGluGlyThrIleProGlnAspPheLysAspGln 2116
    |||||
QY 6589 ATGATATATCATATGTGTATTAAGATATGTTTTTGTGAACGTATATTTCTAATGAT 6648
    |||||
Db 2117 MetPheTyrThrTyrGluAspTyrArgAspPheLeuPheGlyThrAspIleSerLysGly 2136
    |||||
QY 6649 AAA-----AAATTTATCTGAACAAATAGTGAACAAACATTCCTCAATGAAAATTAATAG 6705
    |||||
Db 2137 HisGluLysGluSerAlaLeuGlyLysLysIleAspSerLeuPheLysAsnGluAspGln 2156
    |||||
QY 6706 AAAAACACAGATTAATAAATAAAGATGAAGATTAATGATTTTGGCGCAAAATAA 6765
    |||||
Db 2157 LysSerProSerGlyLysThrProThrGlu-----TrpTrpAsnAspTyrGly 2172
    |||||
QY 6766 AAATTTATTTGGAGAGATGATATATGATTAATCTATCATCTCACAGACGAAACGAA 6825
    |||||
```


[illegible]

```

QY 8923 TATAGAAAAGTGATTAAGAAATTTTAAAAAACTTCTACTCTGCTTCACTGCA 8982
Db 2809 -----Ser 2809
QY 8983 GGACAATTGTTAGTCAAAAATATTAATGCGAAGAGATTGTGCTTTGAGCAATGAA 9042
Db 2810 ProSerLeuLeuLysProSerPhe----- 2817
QY 9043 TATAGTATGACAGATTATTCGATATATTAAGAACTGATATGATGACACTTCATTA 9102
Db 2817 ----- 2817
QY 9103 TCTGAAAAAATTTAAAAAATATTGAAACATCAATGAGCAACGAAAAATCGTAAACA 9162
Db 2817 ----- 2817
QY 9163 TGTGGGAAAAATATAGACATAGATATGCGACGCTATGTTATGTGGATTAATAATTGCT 9222
Db 2817 ----- 2817
QY 9223 ACTTCAAAAAGTACATTAGATGAGATGCTGTCATTAATACCAAGCATGAGAACTAAT 9282
Db 2818 -----PheLeuProLys----- 2821
QY 9283 CAGTTCTTCTGCTTGTTAATTGATGGCGAAGCAAGCATGTAAAGAAAAAACAATGTA 9342
Db 2821 ----- 2821
QY 9343 AGTGAATTCATTAAAAACAATGTCCCTCGTTCAAGAGATTAATTTGAAAGGTCAAGAA 9402
Db 2822 -----LysProLys----- 2824
QY 9403 TTATTAAGACACCTGATGTCGAATGATATTAAGAAATATATAGCTTAATATATG 9462
Db 2825 -----MetLysPheTyrProLysIleGlyIleGly 2834
QY 9463 ATAAAAATTCATGCAATGCAAAATCTTAATATTAATATTAAGCAATTAAGATCAATCTTCA 9522
Db 2835 ValLeuHisProPheIleAsnMet----- 2842
QY 9523 GGTATATATAGCAATTAACCATCTGAAGAAATGTTCACTCATATATTAATCAAAAAAGAT 9582
Db 2842 ----- 2842
QY 9583 TCTCAATGCGCTTTGAGTTAAATGATTAATGAATAGTACAGAACAAAAATATAT 9642
Db 2843 -----ValAlaAspProIleThrIleHisGluThrValAlaLysThrPheAsnAsn 2859
QY 9643 GAAAAATTAATGATTCAAAGAGACTTAATAAAAAATTAATATCTGCTTATATTTGTTGAA 9702
Db 2860 AlaValProGluPhe----- 2864
QY 9703 GATGAAGAACACAAAAATATCTACTAGATGAAATTAATAAGAGAGCAAAAGCT 9762
Db 2865 -----HisIleAsnProAspLysThrAspVal 2873
QY 9763 CGTCTTAAGCACTTAATTTCTTACACCCCATGATGATTTCTTATCAAGCACTTTTA 9822
Db 2874 AlaProPro-----ThrLysAsnIle----- 2880
QY 9823 TTCTCAACACATGAGTAGACAAATATGATCTTAATAATATATTTGAAAGTAGATATC 9882
Db 2881 -----LeuAsnGluValLeuProSerAlaIle 2889
QY 9883 TCTGTTGTTATGATGCGGCTTAGTGGTTGATAGCGCTTCATTTCAATGAGAAAAATTC 9942
Db 2890 ProValGlyIleAlaLeuAlaLeuGlySerIleAlaPheLeuPheLeuLysLysLysThr 2909
QY 9943 AAATCGTGTGAGACTTTGCTGATATACGAATATCCCGAAGAGAGATGGAATGCTT 10002
Db 2910 LysHisProValAspLeuPheSerValIleAsnIleProLysSerAspTyrAspIlePro 2929

```

```

QY 10003 ACCTTGCAATCCAAAAATAGTACATACCATATAGAGATGCTCCATTAAGCCAAA---- 10059
Db 2930 ThrLysLeuSerProAsnAsnGlyTyrIleProTyrThrSerGlyLysTyrAspGlyAsnGly 2949
QY 10060 ACATATATATATATATGGA-----GGAGATACATAGTGGAGCATGAACTAAATATATGCG 10113
Db 2950 ThrPheThrLeuLysGluIleValGlyThrAspSerGlyTyrThrAspHisTyr----- 2967
QY 10114 GACTATATCTCTCTGATATATTAAT-----TCATCCGAAAGTACATGATGAGAAATGATAT 10170
Db 2968 -----SerAspIleThrSerSerGluSerGluTyrGluGluLeuAspIle 2983
QY 10171 AATGATATATATATGACAGTACGATGCTTAATATTAATAACATGTTAGAGTACTAGAA 10230
Db 2984 AsnAspIleTyrHisIleValLeuLeuAsnIleLysThrLeuIleGluValValLeuGlu 3003
QY 10231 CCATCAAAAAGG-----GATATPCCAAGCAT-----GATACACCAAGTAAT 10272
Db 3004 ProSerGlyLysLeuSerGlyAsnThrIleProThrSerGlyLysAsnThrProSerAsp 3023

```

RESULT 4

T28432

Variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
N:Altenente names: erythrocyte membrane binding protein 1 (EMP1)

C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: T28432
R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herfeldt, J.A.; Peterson,

Cell 82, 89-100, 1995

A:Title: The large diverse gene family var encodes proteins involved in cytoadherence

A:Reference number: Z20487; MID:95330813; PMID:7606788

A:Accession: T28432

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3078 <SUX>

A:Cross-references: EMBL:L40608; NID:g886374; PID:g886375; PIDN:AAA75396.1

C:Genetics:

A:Gene: var-1

A:Introns: 2611/3

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
8	48e-165	3501.00	1057	470	934	1490	147
Percent Similarity:	38.65%						
Best Local Similarity:	26.75%						
Query Match:	18.00%						

US-10-087-013-1 (1-10628) x T28432 (1-3078)

```

QY 25 ATATATAAATGGGCAATGCGATCATCATTTAGAGAGAGATGCTTAAGCCCTATTATA 84
Db 4 LeuAlaLysMetGlyProLysGluAlaIleGlyLysAspIleGluAsp----- 20
QY 85 AAAGAAAGTACAAAGTGTGAGAAATGTTTGAACGTTATGCCAAAAATATTAAGCAT 144
Db 21 -----GluSerAlaLysHisMetPheAspArgIleGlyLysAspValTyrAsp 36
QY 145 CCATCAAAA-----TATGCAAAAGAACATGTGATTCGTGAAGGAGTTTGAGCAAGCA 201
Db 37 LysValLysGluGluAlaLysGluArgGlyLysGlyLeuGluIndLysArgLeuSerGluAla 56
QY 202 GAATTCGTGCTGCTCTCTACGCCAGTAATTAAGCATATATTAATTCATATCA 261
Db 57 LysPheGluLysAsnGluSerAspProGluInThrProGluLysP-----Pro 71
QY 262 TGTAAATTTAGATGATAGCAACATCTAATTTACGATGATGATGATGATGATGATGATGAT 321
Db 72 CysAspLeuAspHisLysTyrHisThrAsnVal-----ThrThrAsnValIleAsn 88
QY 322 CTTGGCATGTGTGAGCAACAAACCGATTTGAGAGATGAGAACAACTGTAAGT-----GGA 378
Db 89 ProCysAlaAspArgSerAspValArgPheSerAspIleLysGlyGlyGlyGlyGlyGlyGly 108

```

```

QY 379 AATAAATACGTAATTTTAAAGAAAAAGATGCTATAGCCGTGCGCCACCTAGAGA 438
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 109 AsnArglleYsaSpSerGlnGlnIYAspAsnIYsGIYlaCysAlaIProTYArgArg 128
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 439 CGACATATGCTGTAATAAATCGTCAATATATTAATACCCAAATATTCAT 498
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 129 LeuHisValIYsAspGlnAsnLeuGlnIleIuProIleYsIleThraAsnThHIs 148
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 499 GATTATGGAATAGTACTAGTTCAGCAAAATACAGAGTGAATCAATTTGTAATAT 558
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 149 AsnLeuValIAspValIYsMetAlaIalYpHeGlnIYlGlnIserIleThHInasp 168
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 559 CATCCA---CATAAGAACT-----TCAGACCTGTACTGCTTGGCA 600
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 169 TyPProIYsIYrGlnAlaThrIYrGIYAspSerProSerGlnIleCysThHMetLeuAla 188
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 601 CGAAGTTTTCAGATATAGTGAATGATATGTAAGAAATGATATGTTTAAACCAATGTC 660
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 189 ArgSerPheAlaIAspIleGIYAspIleValIArgIYArgAspLeuTYrLeuGIYAsnPro 208
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 661 CATGAC-----AAAGTAGAAACGGCTCCGAGAGGTTTTCAGAAATA 705
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 209 GlnGlnIleIYsGlnArgIlnGlnIleuGlnIAsnAsnLeuIYsThHlePheGIYIYsIle 228
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 706 CATGAT-----GGATGGAAGATGAAGTAAATGATTCATTCCTGATGATGC 756
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 229 TyrGIYIleuAsnIYsAlaGlnIAlaIArgIYrGIYAsnAsp-----ProGln----- 244
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 757 GGAATTTATTAATTAAGAGAAAGCATGATGTAATGAATGAATTAAGATGGA 816
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 245 -----PhePheIYsLeuArgIuAspTrPTyrThHAlaAsnArgIuThrValITripIYs 262
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 817 GGTATTAACATGAGCAGATATAATCTGATATTTATGCAATCAGAAAGTAAATACA 876
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 263 AlaIleThHcysAsnAla---TrpGIYAsnThHTrYPhe---HisAlaThHcysAsnArg 280
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 877 CCATTAATTTCAATCCATAATCGCGCCATAAACAAGAAAGGTCCTCCCAATTTAGAT 936
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 281 GlnGlnIArgThHleYsGIYrIYcysArgIYsAsnAspAspGlnValIProThHTrYPheAsp 300
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 937 TATGTCCTCATATTTACGTGTTTCGACGATGCGGAGAGAGTCTTCCGAAAAAGA 996
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 TYrValIProGlnIYrLeuArgTrIPheGlnIuTrPAlaGlnAspPheCysArgIYsIYs 320
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 997 AATATTAATGAAGGTCAGAGACCTCTGCTGCT-----AATGACAAAGA 1044
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 321 AsnIYsIYsIleIYsAspValIYsArgIYsncysArgGIYIYsAspIYsGlnAspIYsAsp 340
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1045 CGCTTATATGTAATCAATGACATGATGTTACGACAACTAATTTGAAAAAAGCTATT 1104
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 341 Arg---TYrCysSerArgAsnIYrIYrAspCysGlnIYsThHArgAlaIleGIYIYs 359
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1105 TTGCAATTTGGAATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 1164
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 360 LeuArgIYrGIYIYsGlnIYsIleSerCysLeuTYrAlaCysAsnProTYrValAspTrp 379
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1165 TTAGGGAATCAACAAGAGACATTTTAAACAAAAGAAAAATGAAAAAATACAA 1224
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 380 IleAsnAsnGlnIYsGlnIuPheAspIYsGlnIYsIYsIYsIYsIYsIYsIYsIYs 399
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1225 TCATAT-----TTATCGAAGATACCAATTTGTCAAT 1257
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 400 LysIYrGIYIuAsnGlnIYsIleSerGIYIYsIleArgIuIYsAspAlaGlnIYlThHTr 419
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1258 AATATTAATAGTGA---TATTAACAATTTATGAAAAACTTAAAGAAACCAATAT 1314
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 420 ThHrAsnTrYPAspGIYrGIYIuIYsIYsPheTYrAspGlnIuLeuAsnIYsSerGIYIYr 439
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1315 GCAACTAATGACACTTTTAAATTTAATTAATGAAGAAAGTATGT----- 1362
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 440 ArgThrValIAspIYsPheIleuGlnIYsLeuSerAsnIuIleCysThHTrIYsValIYs 459
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Db 808 AsnTyrGluLysAspLeuAlaAspLeuGluIleHisGluAspGluGlu 827
OY 2185 ---AAAGACAAATATACAAAGACATGGAACATCCCATATGCAACAAACCCG 2241
Db 828 LysGluLysGluAspGluAsnGluCysIleGluGluGluLysAsnPro 847
OY 2242 TGTGTAAACCTGTGGAGCGACAA---CCGACTAAATATTAAGAAATAGCAAA 2298
Db 848 CysSerGluLysSerGluAsnLysArgTyrProValLeuAlaAsnLysValAlaTyrGln 867
OY 2299 TACTTTAAAGAGTCATACGAGCAACGCAATCGTGTCTTCAATTAATGAAGA 2358
Db 868 MetHisHisLysLysAlaLysThrGlnLeuAlaSerArgAlaLysSerAlaLeuArgGly 887
OY 2359 AAGCACACAGAGGTATATATTAACGTGGGTAGAGAAAGACCTTCAGACAAATTA 2418
Db 888 AspIleSerLeuAlaGlnPheLysAsnGlyArgAsnGlySerThrLeuLysGluGlnIle 907
OY 2419 TGTAGAAATATGATTAACATCTTAATCTTAATCTGTGTTTCAATGAGCAGATGAT 2478
Db 908 CysLysIleLeuAsnLysSerLysAsnAspSerArgGlyAsnSerGlyProCysThr 927
OY 2479 GCGAAGCGCAGTGTATGATACAAACAAGTTTGCTAGAACTGAATGGAAAGT 2538
Db 928 GlyLys---AspGlyAspHisGlyGlyValArgMetArgIleGlyThrGlyTyrSerAsn 946
OY 2539 GATCCGACACATCGCTAAAGATCAGAGATGTTATTTATGCTTACAGACAGCAT 2598
Db 947 IleGluGluLysLysGlnThrSerTyrLysAsnValPheLeuProProAlaGlyGlnHis 966
OY 2599 ATATGATACATCAATTTGGAACATTTACAAAGCATGATCACCCACTTAATGATAT 2658
Db 967 MetCysThrSerAsnLeuGluAsnLysVal-----GlySerVal 980
OY 2659 GTT---GATGATTTAGTATATATCTTTTGGGGAGTGTCTCTATACGAAATAT 2715
Db 981 ThrLysAsnAspLysAlaSerHisSerLeuGluLysValGlnLeuAlaLysThr 1000
OY 2716 GAAGCAAAACAAGATATACAGATGATTAAGAAAGAAATTAACCTAAAGGCCCAAGAA 2775
Db 1001 AspAlaAlaIleLysIleLysArgTyrLysAspGlnAsnIle-----Gln 1016
OY 2776 GTACTGACCCA-----AAACACAGACAATCTGTGTGAGCTATACCTACAT 2826
Db 1017 LeuThrAspProIleGlnGlnLysAspGlnLysAlaMetCysArgAlaValAlaTyrSer 1036
OY 2827 TTTCAGATTTAGTGTATATATTCGAGAGAGATCTCGGAAAGAAAGACGGTGCATG 2886
Db 1037 PheAlaAspLeuGluLysPheIleArgGlyArgAspMetTyrAspGluAspLysSer 1056
OY 2887 GTAAGCTGACAGACATTTGGAACACTTTTGTATATACATATACATACCTCAAGGC 2946
Db 1057 ThrAspMetGluThrArgLeuIleThrValPheLysAsnIleLysGluLysHisAspGly 1076
OY 2947 ---AAAGAAATGATTAATGATGAT-----GCCCCAAATTTTAAATGG 2994
Db 1077 IleLysAspAsnProLysTyrThrGlyAspGluSerLysLysProAlaTyrLysLeu 1096
OY 2995 AGGCAAAATGTGGGAGGATATAGCCAAAGTATGGAAGGAGCATGAATGGAATATA 3054
Db 1097 ArgAlaAspTyrPheGluAlaAsnArgHisGlnAlaIleTyrGluAlaMetLysCys----- 1114
OY 3055 AAATATTTGAAGATTAATCGGACACACAAACAAAGTATGTTTCGGATATAGT 3114
Db 1115 -----AlaThrLysGlyIleLeuLys----- 1121
OY 3115 GATCATACACATTTGATGATATATCCCAAAATTAAGATGATGACCGAATGGCA 3174
Db 1122 ProGluMetProValAspTyrIleProGlnArgLeuArgTyrMetThrGluTyrPala 1141
OY 3175 GAATGACTGACAAAGTGCGAAGAAAGAGATGATGATGAAAGGAAAGTGTAAAGAG 3234
|||||

Db 1142 GluTyrTyrCysLysAlaGlnSerGlnGluTyrAspLysLeuLysIleCysAlaAsp 1161
OY 3235 TGTAAAGATTAAGATATGCTCAAGCGCTGACGAAAGACAGTGGTACAGCTTACAGAG 3294
Db 1162 CysMetSerLysGluAspGlyLys---CysThr---GlnGlyAspValAspCysGlyLys 1179
OY 3295 TGCACAGACCTGTATATGATATATATGATATATATGATATGAAAGAAACATGGAAT 3354
Db 1180 CysLysAlaIleCysAspLysTyrLysGluGluIleGluLysTyrAsnGlnGluIntParg 1199
OY 3355 ATATATACAT 3414
Db 1200 LysIleSerAspLysTyrAsnLeuLysLeuLysThrLysThrSerThrAsnPro 1219
OY 3415 GGTATTAAGCTTCACAGTACCTGCCAAATAATCATATAGAC-----AGCAATGTAT 3465
Db 1220 Gly-----ArgThrValLeuGluLysAspAspProAspTyrGlnGlnMetVal 1235
OY 3466 GAATTTTGTGGAAATTAACCAACAAATGCTGGCAAA-----AGTAATMAA 3513
Db 1236 AspPheLeuThrProIleHisLysAlaSerIleAlaIleArgValLeuValLysArgAla 1255
OY 3514 AGTGACTAGTATGATGAAGTGTCTGATGATGATGATGATGATGATGATGATGAT 3573
Db 1256 AlaGlySerProThrGluIleAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1275
OY 3574 GCATATCTCCATGATACAGAAATTTTGAATGATGATGATGATGATGATGATGATGAT 3633
Db 1276 GlyTyrIleHisGlnGluIleGlyTyrGlyCysGlnGlnGlnGlnGlnGlnGln 1295
OY 3634 GAAAA-----AGTGAATGATGAATGAATGAATGAATGAATGAATGAATGAAT 3675
Db 1296 LysLysHisGlyAlaThrSerThrSerThrThrLysGluAsnLysGluTyrThrPheLys 1315
OY 3676 GATTAACACAGACAGCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3735
Db 1316 GlnProProGluTyrAlaThrAlaCysAspLysLysAsnArgSerGlnThrGlu--- 1334
OY 3736 GTACAGATTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAT 3795
Db 1335 -----GluProLysLysLysGluGluLysAsnValGluSerAlaCysLysIleValGlu 1351
OY 3796 GATATCTTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAT 3855
Db 1352 LysIleLeuGluLysLysAsnGlyArgThrThrValGlyLysCysAsnProLysGluSer 1371
OY 3856 AGTAATGATATCCCGATGCGCAATGC---GAAATTAATTAATTAATTAATTAATTAAT 3912
Db 1372 -----TyrProAspTyrAspCysLysAsnAsnIleAspIleSerHisAsp---Gly 1387
OY 3913 GTGTGATGCCCCCTAGACAAAGATGATGCTGATCATTTCTGCAAAATGATGAAGAA 3972
Db 1388 AlaCysMetProProAlaArgGlnLysLeuCysLeuTyrGluIleAlaHisGluSerGln 1407
OY 3973 ATAAAAATTTACATGATACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4032
Db 1408 ThrGluAsnIleLysThrAspAspAsnLeuLysAspAlaPheIleLysThrAlaAlaAla 1427
OY 4033 GAAACATTTCTTCATGATGATATTAATTAAGATAG---GATGCTGAAGAAATGAACTC 4089
Db 1428 GluThrPheLeuSerTyrGlnTyrTyrLysSerLysAsnAspSerGluAlaLysIleLeu 1447
OY 4090 GATTAAGAAATTAAGAAAGCAAAATTCCTCCGATTTTGTGAGATCCATGCTTACACA 4149
Db 1448 AspArg-----GlyLeuIleProSerGlnPheLeuArgSerMetMetTyrThr 1463
OY 4150 TTGTGAGATTTATAGAGATTTTATTTTGAACAGATATATCAAAAGTCAATGAGAGGA 4209
Db 1464 PheGlyAspTyrArgAspIleCysLysAsnThrAspIleSerLysLysGlnAsnAspVal 1483
OY 4210 AGTAATTAAGAAAGCAAAATGATGCTCTTTTCAAAATGCTGACAAATGCTCAAT 4269
Db 1484 AlaLysAlaLysAspLysIleGlyLysPhePheSerLysAspGlySerLysSerProSer 1503
|||||

```
OY 4270 GGAAGAACACGAGAGATGGTGGACAGAACATGATCATGATATGGAAGTATGCTA 4329
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1504 GlyLeuSerArgGlnGlnIuTrpLysThrAsnLysProGluIleTrpLysGlnMetLeu 1533
OY 4330 TGTGCATAGTAAATAATTTGGGCGCAAAAAGATGAT-----TTTACCGAAACTAC 4380
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1524 CysAlaLeuThrLysTrpValThrAspThrAspAsnLysArgLysIleLysAsnAspTyr 1543
OY 4381 GGTTCACACACGTCGAATTTAGTACAAA---AGCACCTTTGGAGATTTGCCAAA 4437
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1544 SerTyrAspLysValAsnGlnSerGlnAsnGlnAsnProSerLeuGlnGlnAlaAla 1563
OY 4438 CGACCCAGATTTTACATGGCTTACCCGAATGGTACGACGATTTGCTATACAGCAAA 4497
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1564 LysProGlnPheLeuArgTrpMetIleGlnTrpGlnLysGlnLysGlnLysGln 1583
OY 4498 AATATTTGAAGATGGCGCAAAAATGTAATCA---AATGACCAATTGAAGTGTAT 4554
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1584 LysLysGlnAsnIleIleLysAspAlaCysAsnGlnLysAsnSerThrGlnGlnCysAsn 1603
OY 4555 ACA-----GAATGTAAATAAGAAATGCGAGACTGCTTAATATATGAAA---AAA 4602
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1604 AspAlaLysHisArgCysAsnGlnAlaCysArgAlaTyrGlnGlnLysGlnLys 1623
OY 4603 AAAAAGAGTGGATTCACCAAGATAATATATACAGATGACGCGACAAAAGATTC 4662
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1624 LysLysGlnPheSerGlnTrpAsnAsnPheValLeuLysAlaAsnValGlnProGln 1643
OY 4663 GATGACACACATGTGTGTAATGTTACAGACTATATGAAAGCAATGCAACAGTTTC 4722
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1644 AspProGlnLysLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 1654
OY 4723 TGTACACAGAAATTTACTCTAGTGTGTGATAGCCTGGAAGTCTCTGTGTACAA 4782
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1655 -----GlyValGlnPro-----LysGln 1660
OY 4783 AGAATATATCAATTTGTAGAAAACAGGCTTACTATGATGCGGACAAATTTGTGGTTC 4842
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1661 GlnAsnGlnLysLeuLeuGlnLys-----CysAspAsnAsnLys---CysSerCys 1676
OY 4843 ACAAAATTTATTTGAAAATGACGACAAATATACTAATTTGAGTAAAGTAAAGCAAA 4902
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1676 ----- 1676
OY 4903 GGATAGTAAAGAGGCAACACAGGTCTATTAAAGTGCGCAAAACAAAGCTTAATAAC 4962
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1676 ----- 1676
OY 4963 TACAATTAATCTGAAGAATTGACTGAAGATGTCTTTTCTCTGTGCTGACTACTATA 5022
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1676 ----- 1676
OY 5023 TGTTCATGCATTGTGATGGCAATTAATACAGATCCAGAGTAAAGATGAAATGGTTG 5082
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1677 -----MetAspGlnLysValLeuSerValSerProLysGlnLys----- 1689
OY 5083 CGAAAAGATGATGAGAGTGGCGGCAACGAGAGTACAAATTTGGCAATTAATCTACAAA 5142
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1690 -----ProPheGlnLysTyr----- 1694
OY 5143 GAAAAAAGAAAAAGAAAAATATAAAAGCTGATGCGCACAAAATTTCTATAGAGTTC 5202
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1695 -----AlaHisLysTyrPro----- 1699
OY 5203 CCGCTTGTACTGCTATGAAATATATAGTTTATGATTAAGATATAAATCTAGATAT 5262
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1699 ----- 1699
OY 5263 GATATATTTGAAGATGAAAAACAAAGACCGAGAAAATTTGAAGAAAATATTTAAACAAA 5322
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1699 ----- 1699
OY 5323 AATGAAATCACTATTGGCAAAAGAGTGAATAGTACTACAGAAAATCCGGTAGTACTGCG 5382
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1699 ----- 1699
OY 5383 CGAAAATTTTCTGGAACGAAAAATAAGATGTTGTGACGCAATGATATCCGGGTAC 5442
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1699 ----- 1699
OY 5443 AAACGTGTAGGAGTATGAGAAATAGTGAATAAGTCAAGAAAGTATGAGATCTAAA 5502
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1699 ----- 1699
OY 5503 AATGTGTCTGTACCTTCAGATGATATATTCATATGGGAAAAATCGCATGAGGT 5562
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1699 ----- 1699
OY 5563 ACTGCTATAGTTCCTTCGATGTTTGGCGAATGGGGTGAAGATTTTGCAAACATAAA 5622
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1699 ----- 1699
OY 5623 GAAAAGAAATTTGAGAAATTTGGTAGGGCGCTGAATGATTAATCTTGTGTATATGAA 5682
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1699 ----- 1699
OY 5683 GATAAAAGAAAGAAATGTACAGATGCGGTACCAATATATAAAATTTATAGTAGTGG 5742
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1700 -----GlnLysCys----- 1702
OY 5743 AAACCAAGTATGAAAAACAATCAAAATATATGTGAGAAATAAGACAAAATATATTC 5802
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1702 ----- 1702
OY 5803 GAGCATCCTGTGGCAAAAGATGCACAGACGCTGCGCAATATTTAGACAAAATTAATA 5862
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1702 ----- 1702
OY 5863 AAAATTTGTGAAAAATTAAGTGAATTTGTAATATATAGTATGAAAGATGTGCCACA 5922
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1703 -----AspLys----- 1704
OY 5923 CAGCATTAATGATGATTAATAGTCAAAATATATGCCCGCATTAAGACGATGAACAAA 5982
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1704 ----- 1704
OY 5983 GAAGTTGAAGAAAGTGAATTTGCAAGTGTCCACAGAGTCCACACGATGTACAGAGGAA 6042
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1705 -----TyrGlnGlnLys-----HisValProSerIleProPro----- 1715
OY 6043 ACACCGTACCAAGGATATCACTGATATCAAAAGCAGCGCATCGAAAAAGACGAAA 6102
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1715 ----- 1715
OY 6103 ACAGCGCGCTTACAAAACAGCCGAAAAAGTGAATAATCTTAACACAGAAATCGCAGCA 6162
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1716 -----ProProProProValGlnProGln----- 1723
OY 6163 CAACACAGAACCCGACGACGACGACAAACAAACAGAAACGACATCAACAGCAACACA 6222
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1724 -----ProGlnAlaProThr----- 1728
OY 6223 ACAGAAATCTGACGTGGGACACAAATGTAAGGCAATCTTTCGATAAACCAGATAGCAGG 6282
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1729 ValThrValAspValCysSerIleValLysThrLeuPheLys-----AspThrAsn 1745
OY 6283 GGTGGAATGAGGGGTGTATATCCAAAAGCTATAGCAATAT-----CTTAATGGCGGT 6336
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1746 AsnProSerAspAlaCysGlnLysLys-----TyrGlnLysThrAlaProSerSerTrpLys 1764
OY 6337 TGAATTT-----GTAGTAAAGTATAAGAAAAT 6363
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1765 CysIleProSerAspThrLysSerGlnAlaGlnAlaThrThrLysSerGlnLysSerGln 1784
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 6364 GAAAATGCAATATGATATGCTCTCTAGAGCAAAAAATATATATATATATATATATAT 6423
```

```

Db 1785 SerGlySerIleCysIleProProArgArgArgLeuValGlyLysLeuGln --- 1803
QY 6424 TTAATATGAACT -----GAAATTAAGCGTAC 6453
Db 1804 ---GluThrIleAlaThrAlaLeuProGlnGlyGlyAlaAlaProSerHisSerAlaGala 1822
QY 6454 AATGATATTAAGAGCGCTTTATTAATGTGACAGCANTAGAACTCAATTTTGTGGTTA 6513
Db 1823 AspAspLeuArgAsnAlaPheIleGlnSerAlaAlaIleGluThrPhePheLeuTrpAsp 1842
QY 6514 AATATATATATGAA---AATCTGCAGCAGAAATGAA----- 6549
Db 1843 ArgTyrLysGluGlnLysLysProGlnGlyAspGlySerGlnAlaLeuSerGlnLeu 1862
QY 6550 -----TTGCAAAATGCAACA 6564
Db 1863 ThrSerThrTyrSerAspAspGluLysAspProProAspLysLeuGlnAsnGlyLys 1882
QY 6565 ATTCCGAGATGAATTTAAAGAAATATGATATTAACATATGCGTATTAAGATATGCTT 6624
Db 1883 IleProProAspPheLeuArgLeuMetPheTyrThrLeuGlnLysAspTyrArgAspIleLeu 1902
QY 6625 TTGGAACTGATATTTCTATGATTAATAAATAATATACTTAACAAATAGTGTAAACACC 6684
Db 1903 ValHisGlyLysAsnThrSerAspSerGlyAsnThrAsnGlySerAsnAsnAsnIle 1922
QY 6685 ATTCTCAATGAATTAATAAGAAATATGATATTAACATATGCGTATTAAGATATGCTT 6744
Db 1923 ValLeuGlnLysSerGlyAsnLysGluAspMetGlnLysIleGlnLysIleGluGln 1942
QY 6745 ATA----- 6747
Db 1943 IleLeuProLysAsnGlyGlyThrProLeuValProLysSerSerAlaGlnThrProAsp 1962
QY 6748 ---TTTGGAGAGAAATAAATAATTTATTTGGAGAGAGATGATATGATTAACCTAT 6804
Db 1963 LysTrpAsnGlnHisAlaGlnSerIleTrpLysGlyMetIleCysAlaLeuThrTyr 1982
QY 6805 -----CATCTCAGACAGAAAGAAAGAAAGAAAGAAATTTAGA 6840
Db 1983 ThrGluLysAsnProAspThrSerAlaArgLysAspGlnLysIleGluLysAspAsp 2002
QY 6841 GATATTTAC----- 6849
Db 2003 GluValTyrGluLysPhePheGlySerThrAlaAspLysHisGlyThrAlaSerThrPro 2022
QY 6850 -----CAGTACAAT---GACATGACCAACGTG----- 6873
Db 2023 ThrGlyThrTyrTrpThrGlnTyrAspTyrGluLysValLysLeuGlnLysAspThrSerGly 2042
QY 6874 -----ACGCTTCCCTGAGAGATTTGTAATAAGAGCC 6906
Db 2043 AlaLysThrProSerAlaSerSerAspThrProLeuLeuSerAspPheValIleuArgPro 2062
QY 6907 CAATTTTGGATGCGTTACACAGAAATGGCAGAAAGATTTGTAATAAGAGAGAGAAAG 6966
Db 2063 ProTyrPheArgTyrLeuGlnGlnTyrPGLysAsnMetLysLysArgLysHisLys 2082
QY 6967 TTGTTAAATTTGAGCGGCGCTGTAAGGATATGAG----- 7002
Db 2083 LeuAlaGlnIleLysHisGlnCysLysValGlnGlnLysGlnGlySerAlaArgGly 2102
QY 7003 -----TGTAAAT-----GGTAGTAAT 7017
Db 2103 GlyIleThrArgLysThrSerGlyAspGlyGluAlaCysAsnGlnMetLeuProLysAsn 2122
QY 7018 GAGGATAGACACAAGA-----TGTCAGAGCGCGTGTGAACATATCAA 7062
Db 2123 AspGlyThrValProAspLeuGlnLysProSerCysAlaLysProCysSerTyrArg 2142
QY 7063 AATTTTATTAAGAAAGTGAATGAAATGAAAGCAAAAGAAAGTCAAAAGAT 7122

```

```

Db 2143 LysTrpIleGlnSerLysGlyLysGluPheGlnLysGlnLysAlaTyrGlnGln 2162
QY 7123 AAAGAT-----GGCAAAAAGATATAGGATATATCTTACTGAAAGACATA 7170
Db 2163 LysAspLysCysValAsnGlnLysSerAsnLysHisAsp----- 2174
QY 7171 GAGNAGCAACATGTGCTCATGAAATTTAAACATGAAATTAAGAAATATGTGCAAT 7230
Db 2175 -----Asn 2175
QY 7231 AAGGATGTCTTGTATGCAAAAAACCTTTCACAACTACCAAAAAACAAACAATCA 7290
Db 2176 GlyPheCysGluThrLeu-----Thr 2182
QY 7291 CAATCATCCGATGCTATGATATATGACCAATCCGATGATATGTCTGTAAGATTTAAC 7350
Db 2183 ThrSerSerLysAlaLysAspPheLeuLysThrLeuGly----- 2195
QY 7351 AAGTGTGAGTGTCTGCAACTTTCAAAAAGGATCTATGAT----- 7392
Db 2196 ---ProCysLysProAsnAsnValGlnGlyLysThrIlePheAspAspLysThr 2213
QY 7393 -----CATCAAAAAAATTTACTGAACT-----AAATACCTATGATTTGTGTAG 7440
Db 2214 PheLysHisThrLysAsp---CysAspProCysLeuLysPheSerValAsnCys----- 2230
QY 7441 AAAGCAGCATATTTATTTATCTAAAGACAGCAAAATTAATGATATTTACTTGAAGAA 7500
Db 2230 ----- 2230
QY 7501 AATTTATACCTATTTAGTGTACAAAGAAAGAAAGAAAGTAAATAGTGTGACTAATAT 7560
Db 2231 -----LysLysAspGlnCysAspAsnSer----- 2238
QY 7561 AATCCTTGCGATCCTAAGAAACCTTATGCAACGATTAATATATAGAGAGAAACCT 7620
Db 2239 -----LysGlyThrAsp 2242
QY 7621 TGTGAAATATAGAGAAAGAAATCGTTTAAAGTAGATTTATGAAAGAAATGTACAAAT 7680
Db 2243 CysArgAsn-----LysAsn 2247
QY 7681 TCAAGTTCATCAGAGAAAGAAAGAAAGATGTGTACTCTCCAGAAAGAAACATATGTGC 7740
Db 2248 Ser----- 2248
QY 7741 TTAAGCAATTTAGATGAATTAATTAATGAAGACTTAAGATAGTATTAATCTCTAATA 7800
Db 2249 -----IleAspAlaThrAspIleGlnAsnGlyValAspSerThrValLeuGlnMet 2265
QY 7801 ATGGTTCGTGAACTGCAGCAAAATGAAGATAGACATAATAAATACTTCAACTCA--- 7857
Db 2266 ArgValSerAlaAspSerLysSer-----GlyPheAsnGlyAsp 2278
QY 7858 -----GAGNAGCGGTGCGCAATGCAATCCAAATATGTATACTATGAATATAGTTCGCT 7911
Db 2279 GlyLeuGlnLysAsnLys----- 2284
QY 7912 GATCTGGGTGACATAGTATGAGNAGCAAGATATGTACGAATTTGGTGTACTTACTCC 7971
Db 2285 -----ArgGlyAlaGly----- 2288
QY 7972 GTAGAAATTAATAATTAATTAAGGTTTGAATACATATATGAAATAAGAGAAATTAATAAT 8031
Db 2289 -----IlePheGlnGlyLe----- 2293
QY 8032 AAAGGTGAATAATAATCAACGATGTACAAACGTTTCTGTTCTGTGTGGATGCTAAT 8091
Db 2293 ----- 2293
QY 8092 AGAAAGATATTTGAAAGCA-----ATGACGTGCAAGACCGAA 8133
Db 2294 ArgLysAspGlnTyrPheCysArgAsnValCysGlyTyrValValCysLys---ProGln 2312

```



```
OY 8134 GATCAAAACCTTTTGAAGAAAGAGATGATGATTGAAACCATTAATACAA 8193
Db 2313 AsnValasnGlyGluAlaLysGlyLys----- 2321
OY 8194 GATAAGTGTGACATAAGAGACATCCCTGTCATATATACCTCAACGTTTGA 8253
Db 2322 ----- 2328
OY 8254 TGGATGACGTAGTGTGATATATTATTGAACACTGATGAGAAATGGAATAATT 8313
Db 2329 LeuValLysArgTyrPheGluTyrPhe-----PheGluAspTyrAsnLysIle 2344
OY 8314 AAAAAATCATGTGATCTCTTAAACATCTGACAGATGACAGATGATGATGAATAAT 8373
Db 2345 LysHis-----LysIleSerHisArgIleLysAsn-----Gly 2355
OY 8374 AAGTGTACAGCTGTAAAGACAGATGTCAAGATATTAATAATTTTGTCTTAAATGAAA 8433
Db 2356 GluIleSerProCys-----IleLysAsnCysValGluLysTyr--- 2368
OY 8434 TCTCTATTCGATATACATCAATCAATTAATACAAAGAAATTTGATGACAAACCATATATACA 8493
Db 2369 -----ValAspGlnLysArgLysGluTyrPheGluIleThrGluArg----- 2382
OY 8494 AAAATCTACTATATGATCATGTTCACAAATTTGTCAAAAGTTGAAAACCTTTAAAGT 8553
Db 2383 ----- 8559
OY 8554 GAATGTTCTGTGAGAGCTTTCTGATATCTCATGAACAAAGTAAGTGTTCATTAAT 8613
Db 2386 Gln-----Tyr 2387
OY 8614 AAATTTAAGAAATATAGTGTCTTCCATATACAGACATATGCTTTGAGAAACACACA 8673
Db 2388 Lys-----AsnAspAsnSerAspAspAsnValArgSerPheLeuGluIleThrLeuIlePro 2406
OY 8674 AAA-----AGTTATTAAGAGCT 8691
Db 2407 GlnIleThrAspAlaAsnAlaLysAsnLysValIleLysLeuSerLysPheLysAsnSer 2426
OY 8692 TGCAGTGTACACTACCTCTTAAGAAATCCATGATTAATGTCTCAACGCAAAACAAA 8751
Db 2427 CysGlyCysSerAlaSerAla-----AsnGluLysAsnLys 2438
OY 8752 GATGATGTAAAGAAATTAACAACCTTTACTCTGCTGAGAAATGATTAATATAT 8811
Db 2439 AsnGly----- 2440
OY 8812 CTTCATTAATGACACATACCTGTCCTTAATAGTTCAGATGATTAACAAGGTATTTG 8871
Db 2440 ----- 2440
OY 8872 ATTCCTCAAGAGAGACATTTATGTACAGACCTATCACTGCATTAATTAAGAAA 8931
Db 2440 ----- 2440
OY 8932 GGTGATTAAGAAATTTAAAAAAACCTCTTACTCTGCTTCAAGTCAAGACAAATTG 8991
Db 2440 ----- 2440
OY 8992 TTAGTCAAAAATATTAATCGAAGAGAGTGTGCTTTGAGGCAATGAATATATATTAT 9051
Db 2440 ----- 2440
OY 9052 GCAGATTATTCGATATTAATTAAGAACATGATATGATGACACTTCTTATTCGAAAAA 9111
Db 2441 ---GluTyrLysAspAlaIle-----AspCysMetLeuLysLysLeuLysAspLys 2456
OY 9112 ATTAATAAAATATTTGAAACATCAATGAAGAACACGAAATCGTAAACATGTGTGGAA 9171
Db 2457 Ile----- 2457
```

```
OY 9172 AATAATAGACGTACATATGCGACGCTATGTTATGTGATATATAAATTCCTCAACAAA 9231
Db 2457 ----- 2457
OY 9232 GTAACATTAGATGAGAGATGCTCAATTACCAAGATGAGAAACTAATCAGTTTCTT 9291
Db 2458 -----GlyIucys----- 2460
OY 9292 CGTTGTTAATTGATGGCGAAAGACAGCATGTAAAGAAAAGAACATGATGATTCGA 9351
Db 2461 -----GluLysLysHisIleGlnThrSer 2468
OY 9352 TTTAAAACAAAATGT-----CCTCGTTCAAGAGATATATTTGAACG 9396
Db 2469 ---AspThrGluCysSerAspThrProGlnProGlnThrLeuGluAsp----- 2483
OY 9397 TCAGATTTATTAAGACACACCTGATGTCAGATGATTAATTAAGAAATATATAGCTTGAAT 9456
Db 2484 ---GluThrLeu-----AspAspAlaIleGlnThrGluAlaLysLys 2497
OY 9457 ATATGATTAATAAATACATGGAATAATCAATTAATTAATTAATTAATTAATTAATCA 9516
Db 2498 AsnMetLeuProLysIleCysGluAsnVal---LeuLysThrAlaGlnGluLysPheGlu 2516
OY 9517 TCTCAGTAAATATAGACAAATTAACCATCTGAGAGAAATGTTCAGTCAATATTAATCA 9576
Db 2517 ---GlyGlyCysValProAlaGluAsnSerGluLysProAla-----Ala 2530
OY 9577 AAGATTCGATCGCGCTTTGGAGTTAAATATTAATTAATTAATTAATTAATTAATTAAT 9636
Db 2531 ThrAspSer----- 2533
OY 9637 AATAATGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 9696
Db 2534 GlyLysGluThrProGluGlnThrProValLeuLys----- 2545
OY 9697 GTTGAAGATGAACACACAAAAATCATGTAAGTGAATTAATTAATTAATTAATTAATTAAT 9756
Db 2546 -----ProGluGluAla 2550
OY 9757 ACAGTGTCTTAAGACACCTATTTCTTACACCCATGAGATTCTTCTATCAACA 9816
Db 2551 ValProGluProPro-----ProProProProGluLysAlaProAla 2565
OY 9817 CTTTATTCACACATACGACATGACATATGATGCTTAATAAT----- 9861
Db 2566 Pro-----IleProGlnProGlnProProProProProProProProProProProPro 9871
OY 9862 -----GATTAATGAAAAAGTAGATCTCTGTTATTTATGATGCGCTATAGT 9909
Db 2580 LeuAspAsnProHisValLeuThrAlaLeuValThrSerThrLeuAlaTyrSerValGly 2599
OY 9910 TTG-----ATAGGCTTCATTCATGAGAAATTAATTAATTAATTAATTAATTAATTAAT 9954
Db 2600 IleGlyPheAlaThrPheThrTyrPheThrLeuLysLysLysLysLysSerSerValGly 2619
OY 9955 GACTGTGGGTATCTGAATATCCGCAAGAGAGATGAGAAATGCTACCTGGAATCC 10014
Db 2620 AsnLeuPheGlnIleLeuGlnIleProLysSerAspTyrAspIleProThrLysLeuSer 2639
OY 10015 AAAATAGTATACATACCATATAGAGTGCATATTAAGGCAAAACATATATATATATATG 10074
Db 2640 ProAsnArgTyrIleProLysTyrThrSerGlyLysTyrArgLysArgTyrIleTyrLeu 2659
OY 10075 GAAGGAGATCT-----AGTGAAGATGAAGATTAATTAATTAATTAATTAATTAATTAAT 10125
Db 2660 GluGlyAspSerGlyThrAspSerGlyTyrThrAspHisTyr----- 2673
OY 10126 TCTGATATTAATCTATCCAGAAAGTGAATGAAGAAATGGAATTAATTAATTAATTAATGA 10185
Db 2674 SerAspIleThrSerSerGluTyrGluGluMetAspIleAsnAspIleTyrVal 2693
OY 10186 CCAGGTAGTCTTAATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10245
```

```

Db      2694   ProGlySerProLysTyrLysThrIleuLeuValIleuGluProSerGlyAsnAsn    2713
          |||||
Oy      10246  ATCCAACTGAT-----                                     10257
          ::::

Db      2714   ThrThraIaSerGlyAsnAsnThrThraIaSerGlyAsnAsnThrThraIaSerGlyLys    2733
          |||||

Oy      10258  GATACACAAGr-----AATGGATACCACCAAGTCACGAATAGATT    10296
          :|||||

Db      2734   AsnThrProSerAspThrGlnAsnAspIleGlnAsnAspGlyIleProSerSerLysIle    2753
          :|||||

Oy      10297  ATTGATGTGCATGATGAATGACTGAAACATGATTTTGATTCATATTTA-----    10347
          |||||

Db      2754   ThrAspAsnGluTrpAsnGlnIleuLysAspGluPheIleSerGlnTyrIleuInserGlu    2773
          |||||

Oy      10348  CCAATACAGAACAACAAATAATTAATTCAAAAGTGCAGATATTCCAATGATACAGAACCT    10407
          |||||

Db      2774   ProAsnThrGluPro-----AsnMetLeuGlyTyrAsnValAspAsnAsnThrHisPro    2791
          |||||

Oy      10408  AATACTTATATTCGATTAATCCTGACAGAAAAACCTTTATATATTCATCATGATAG    10467
          |||||

Db      2792   ThrThrSerHis--HisAsnValGluGluIulysProPheIleMetSerIleHisAspArg    2810
          |||||

Oy      10468  GATTTATATCTACTGGGAAAGAAATTAGTAT-----                10497
          |||||

Db      2811   AsnLeuPheSerGlyGluGluTyrAsnTyrAspMetPheAsnSerGlyAsnAsnProIle    2830
          |||||

Oy      10498  AATATTATATGAGTAGCAAT-----ACATAATATCATATTCCAAAG    10539
          |||||

Db      2831   AsnIleSerAspSerThrAsnSerMetAspSerLeuThrSerAsnAsnIleSerProTyr    2850
          |||||

Oy      10540  AATGCTGAATATGATTTCTTATAGAGCTATAGATTAAATTAATGATTCACCTAGTGCGTA    10599
          |||||

Db      2851   AsnAspLysAsnAspLeuTyrSerGlyIleAspLeuIleAsnAspAlaLeu--SerGlyA    2870
          |||||

Oy      10600  AACCATATGATATATATGATGAAGTATG    10628
          |||||

Db      2870   snHisIleAspIleTyrAspGluMetLeu    2879
          |||||

```

RESULT 5
 T18378
 Variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
 N:Alternate names: erythrocyte membrane binding protein 1 (EM1)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T18378
 R:Bauch, D.I.; Pasloske, B.L.; Singh, H.B.; Bl, X.; Ma, X.C.; Feldman, M.; Taraschi, T
 S:11 82, 77-87, 1995
 Title: Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and
 Reference number: Z18925; MUID:9530812; PMID:7541722
 A:Accession: T18378
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2924 <BAR>
 A:Cross-references: EMBL:U27338; NID:g914918; PID:g914919; PIDN:AA60251.1
 C:Genetics:
 A:Gene: EMP1
 A:Introns: 2476/3

 Alignment Scores:
 Pred. No.:

Score:	9.54e-148	Length:	2924
Percent Similarity:	3154.00	Matches:	1013
Best Local Similarity:	38.74%	Conservative:	453
Query Match:	26.77%	Mismatches:	963
	16.21%	Indels:	1356
	2	Gaps:	150

 US-10-087-013-1 (1-10628) x T18378 (1-2924)

```

0y  97  AAAGTGCAGCAATGTTTGCAGCGTTATCCAAAATATA----- 138
    |||  |||:::|::|  |||::|::|  ::  ::
Db  13  lysAspAlaIaYsHISAlaIeuAspArgIleclYlucIuIaValYrLYsGluYsValGlu 32

```

OY	139	AGACATCCATCAAAATTTGCAAAAGAAACAATGTCGATCTGCTTAAGAGGGATTTGCAGAA	198
Db	33	AsnAspAlaGluTyrTyrLysLys-----AlaLeuLysGlyAsnLeuGlnGlu	48
OY	199	GCAGAAATTCGTGGTGGCTCCTTACGCCAGTAAATAAACATAATTAATTAATCAAT	258
Db	49	AlaLysGlyLeuGlyGluLeuAlaSerSer-----Proasn	60
OY	259	CCATGTAATTTAGATCATTAAGAACTACTAAAT--TTCAGGTATGATGATGTAATTTG	315
Db	61	ProCysLysLeuValGlnAspTyrTyrAsnAsnArgLeuLys-----ArgLys	76
OY	316	AGACATCTGTCGCCATGGTAAAGACAAACCCGATTTGGATGAAGATGAACAAATCTGAATGT	375
Db	77	ArgTyrProCysAlaAsnAlaGlnThrValArgPheSerAspGluTyrGlyGlnCys	96
OY	376	GGA---AATAAAATACGTATTTAATAAGAAAAAAATGATGCPATA---GCCGTGCCGCA	429
Db	97	ThrPheAsnArgIleLysAspSerGluAsnAsnAspSerIleGlyValAcysAlaPro	116
OY	430	CCTAAGAAAGACATATGTCGATTAATAAACTTGGAGGCTCTAAATGATATAATTAATCCCA	489
Db	117	TyrAlaArgLeuHisLeuCysAspTyrAsnLeuGlnLysMetGlyLysThrSerThrThr	136
OY	490	AATATTCATGATTTATTTGCGAAATGTACTAGTTTACGCAAAATACGAAAGCTGAATCAAT	549
Db	137	Lys---HisAspLeuLeuAspValCysMetAlaAlaLysTyrGlnGlyAspSerIle	155
OY	550	GTTAATATCAT---CCACATAAAGCAACT-----TCCAGACCTGTACT	591
Db	156	LysThrHisTyrThrLysHisGluLeuThrAsnProAspThrLysSerGlnLeuCysThr	175
OY	592	GCTCTTGACAGCAAGTTTGCAGATATAGCTGATATTGTTAAGAGGAATGAATATGTT---	648
Db	176	IleLeuAlaArgSerPheAlaAspIleGlyAspIleValArgGlyLysAspLeuTyrLeu	195
OY	649	-----AAACCAAAATGTCATGACGAAAGTGAACAGCGGTCCGAAAGGTT	693
Db	196	GlyTyrAspAspLysGlnLysAspArgLysValLysLeuGlnAsnAsnLeuIleGluLys	215
OY	694	TTCAAGAAATACATGATGGAATGGAAGATGAAGTAAATAATGATTAACATCCTGATGGA	753
Db	216	PheLysLysIleHisGluAsnLeuGlyThrGlnAspAlaLysAspHisTyrLysLysAsp	235
OY	754	TCTGGAATATTAATTAATTAAGAGAAGCATGGTGGAAATGTANATAAGAAATTAATGATG	813
Db	236	GluGluAsnTyrTyrGlnLeuArgGlnAspTrpThrAlaAsnArgSerThrValTrp	255
OY	814	GAACTATACATGTGATGCATCATATAAATCGATATTTATTCATCATCAGAAAGTAAT	873
Db	256	LysAlaIleThrCysHisAlaGlnLysSerAspLysTyrPheArgLysThrCysCysSer	275
OY	874	AACACATTAATTTCAATTCCT---AAATGGCGGCATAAACAGGAAG---	918
Db	276	GlyLysIleThrPheAspAspLysCysAspGlyCysLysAspGluGlnGlyLysAsnGluThrAsn	295
OY	919	---GTTCCTACCAATTTAGATTAATATGCCCTCAATATTTACGTTGGTTCACCAAGGGA	975
Db	296	GluValProThrTyrThrAspTyrValProGlnTyrLeuAlaArgTrpPheGluGluIleThrAla	315
OY	976	GAAAGATTTTCCGAAAAAGAAATATTAAATTGAAGAAAGCTCAAGAGCTCTGTCTGAAT	1035
Db	316	GluAspPheCysArgLysAlaGlyLysLysIleGlnAsnAlaIleLysAsnGlyArgGly	335
OY	1036	GACAA-----GAAGCCTAATATTGTAGTCATTAATGAGCAATGATGTGACAACTAAT	1089
Db	336	GluLysGlyAsnGluArg---TyrCysAspLeuAsnGlyTyrAsnGlyGluIleThrAla	354
OY	1090	TGGAAGAAAGAGTATTTGCAATTTGGATATAAGCTAGTACGACTGTGCTGCTCAAAAGCAA	1149
Db	355	ArgGlyAlaGluIlePheValLysGlyLysAspCysHisLysCysSerValAlaCysAsp	374
OY	1150	GTTTTGAAGTTTGGTTAGGAATCAACAAGACATTTAAACAAACAAAGAAAAATAT	1209

```

Db 375 ArgPheValIysTrpIleAspAsnGlnArgLysGlnPheAspLysGlnLysLysLysTyr 394
OY 1210 GAAAAGAAAACAA-----TCATATTATGCAAGCATTAACAAATTTGTCAT 1257
Db 395 AspIleGlnIleAsnLysThrHisGlyThrThrIleThrGlnGlnLysIleAsn 414
OY 1258 AATATTAACTGATATTATTAACAAATTTATGAAAAACTTGAAGAAAGCAATATGCA 1317
Db 415 AsnLeu-----TyrValGlyHisPheTyrLysIleLeuLysLys-----TyrTyrPro 430
OY 1318 ACTAATGACACTTTTAAATTTTACTAAATGAAGAAAGATTTGTA-----GGAGA 1371
Db 431 ThrValAspLysSerLeuGlnLysLeuAsnAspGlnIleLecysLysLysProProAsn 450
OY 1372 TTACACAGAGAAAAGATATTACTTTTACTAACAGTCGCTGATGCAAGGATATTAT 1431
Db 451 ValGlyAsnGlnLysAlaSerThrValAspPheAsnAsnGlnValAsnThrThrPheSer 470
OY 1432 CGTTGCAATATTGCCAGTGTGTCGCCGACTGCGGGGTCAATGTGAT----- 1479
Db 471 HisThrThrTyrCysGlnAlaCysProTyrCysGlyValGlnLysGlnLysAsnGlyGly 490
OY 1480 GGTATATAATACACACACAAATCA--GATAATGATCGTGAAGCTGTAATATATGAAGAC 1536
Db 491 GlyTyrLysAlaLysGlnLysSerCysAlaLysLysGlnLysGlnLysIlePheAsnLysGln 510
OY 1537 TATTAACCTCCACGAGGGGTGAAGCCTACTAATATCATCTGCTTATAGTGTAAATGA 1596
Db 511 -----AsnSerThrAspIleLysIleLeuThrProGlnLysGly 523
OY 1597 GAGGTGATATTACAAAAATATGAAAAATTTTGTGAACGCTCAACTATTACAAAGAT 1656
Db 524 ArgSerLysThrLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 542
OY 1657 AAAAATATCAAAAATGGAATGCTATTATTAAGGATGAATAAATATATGATGTAACAG 1716
Db 543 -----AsnAspIleTyrLysCysHisTyrAspAspAsnGly-----Thr 555
OY 1717 GAACAAAATACGCAATCAATATGAT----- 1743
Db 556 AspAspGlnThrAspAspSerAsnAspCysValLeuGlyAspTyrPylAsnLeuThrLys 575
OY 1744 AATCCTAATATATATCATTTATTTTGAATTTGGTATGATATATATTAAG 1803
Db 576 GluAspLysIleLeuSerLysAsnAlaPhePheTyrPylHisAspLeuLys 595
OY 1804 GATCTATTAACTGGAAGCAAACTTAAACTGTATTAAT--AATCAACACGACAT 1860
Db 596 AspSerIleLysTyrPylArgAspGlnHisGlyArgCysIleAsnLysAspLysGlyThr 615
OY 1861 TGTATTGATGAATGAAGAAATGCTTAATGTTTGAACAGTGGTTAAACAAAAGAA 1920
Db 616 CysIleLysGlyCysAsnLysLysCysIleCysPheGlnLysTyrValGlnLysLys 635
OY 1921 GAAGATGAATAGTATTAAGAACTGTTCACAAAAAAGAAATATCAGACATCGAT 1980
Db 636 ThrGlnTyrPylLysIleLysAspHisPheArgLysGlnLysAspIleProLysAspTyr 655
OY 1981 TATATATATTAATTAATCTTTTGAAGGTATATTTTAAAGTTATGATTAACCTGAC 2040
Db 655 ----- 655
OY 2041 AAAGATGAAGCAAAATGGAAGAACTTATGCAAAATATAAAAAGAAAAAAGAGTTT 2100
Db 655 ----- 655
OY 2101 TCCAATTTGAAAATATATAGGAGACTATTAGAGATGCAATA-----GAAGCTGTGTTA 2154
Db 656 -----ThrHisAspAspPheLeuGlnThrLeuLeuMetLysAspLeuLeu 671
OY 2155 GATCACTTAAAGAAACCTGCCACGATATGTAAGCAATAATACAAAGAGCATGTGTA 2214
Db 2214 ----- 2214

Db 672 GluIleIleGlnAspThrTyr-----GlyAspAlaAsnGln----- 683
OY 2215 ACATCCCATTAATCAACAAACAAACCCGTCGTAAACCTGCTGAGCAGCAACCCACT 2274
Db 683 ----- 683
OY 2275 AAAAATTAATAAGAAATATGACAAATCTTTAAAGAGATGATACGAGAGACGAAAT 2334
Db 684 -----IleLysArgIle-----GluAlaLeuLeuGlnIleAla----- 694
OY 2335 CGTGTCTTCATTAATTAAGAAAGCAACAGCAGAGATATATTAATTAACGTGGGTAG 2394
Db 695 ---GlyValGlyGlyLysPheAlaAlaLeuAlaGlyLeuTyrThrLysGlyPheVal 713
OY 2395 AGAAAGACTTCAAGCAATTTATGATTAATTAATTAATTAATTAATTAATTAAT 2454
Db 714 AlaGlu-----LysAspThrThrIleAspLysLeuLeuGlnHisGlnLys----- 729
OY 2455 GGTTTTCAATGACCATGATGTCGCAAGCAGCAGGATGATGATATCAAAAGATTT 2514
Db 729 ----- 729
OY 2515 GTCTGAGAACATGAATGGAGAGTGCAGACATCGTAAGATCAGCAGATGTT 2574
Db 730 -----GluAlaAspLysCysLeuLysThrHisThrAsp 741
OY 2575 ATTATGCCCTCCTAGAAGACGACATATATGATCATCAATTTGGAACATTTCAACAGGAT 2634
Db 742 ThrCysProProGln-----Glu 747
OY 2635 GATCAACCCATTAAGTAATATGTTGATGATTTAGTTAATTCCTTTTGGGGAT 2694
Db 748 AspArgSerVal----- 751
OY 2695 GTTCTTCTATCAGCAAAATATGAGCAAAACAGATTAATGCAATCTATTAAGAAAGAT 2754
Db 752 -----AlaArgSerGlnLysThrVal----- 759
OY 2755 AACTAAGAGGCCCCAAGAAAGTAACTGACCCCAAAACACGACATCTGTCGAGCT 2814
Db 760 -----ProSerProAlaAspProLysAlaThrGlnGluVal----- 772
OY 2815 ATAGCTACGTTTTCAGATATAGTGATTAATTCGAGAAAGATCTCTGGAAAGA 2874
Db 772 ----- 772
OY 2875 AACGGTGACATGTAACCTCAAGGACATTTGCAAACTGTTTGGTAATATACATPAAG 2934
Db 772 ----- 772
OY 2935 TCACCTCAAGGCAAGAAATGATTAATATATATGATGATGCCCCCAAAATATTAAATTTG 2994
Db 772 ----- 772
OY 2995 AGGAAAATTTGTTGGAACTTAATAGCCAAAGCTATGGAAGCCATGAATGTATATA 3054
Db 773 -----AspAlaAsnAlaSer----- 777
OY 3055 AATATTGAAGCATTAATGCGGACACCAATCAACACAAAGTAGTTATTCGCGATATAGT 3114
Db 778 -----Ser 778
OY 3115 GATCATCAACCATTTGATGATATATATCCCAAAATTAATAGATGATGACCGAATGGCA 3174
Db 779 AspAspGln-----AspAspPheGlnGln----- 787
OY 3175 GAATGTACTGCAAGGTGCGAAAAAGAGATATGTTGAAGAGAGATGAAGAG 3234
Db 788 -----GluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 801
OY 3235 TGTAGATTAAGATTAATGCTCAAGGCTGTAAGAA-----GAGAGTGTACAGGT 3285
Db 802 ValGlnGlnGlnLysThrAspGlnSerAlaThrGlnAlaValAlaProSerProGly 821
```

QY 3286 TGTACGAGTCGACAGAGCTTGTATGAATATATGATATATAGGATTATGGAAGAA 3345
DB 822 ThrThrGln----- 824
QY 3346 CAATGGAATATATATATGAGTAAATACAAAGATTTACATGACAGACAAATCTCTGT 3405
DB 825 ----- 838
QY 3406 AGTAAATAGTATGTAAGCTCCAGTACGCCAAAATCATATAGACAGAAATGTATT 3465
DB 839 CysSerIle-----ValAspLysAlaLeuLys 847
QY 3466 GAATTTTTCGCAATTTATCCAAACAAATGTCGCAAAAGTAAATTAAGTGTACTAGT 3525
DB 848 GlysLeuAspAspAlaCysThrLeuLysTyrGlyLysThr-----AlaProThrSer 865
QY 3526 GATGAAGCTCTGTCATTTGTACTAACACAGCATATGAAATGTTGAGACATATCTCAT 3585
DB 866 TrpLysCysIleProSerGlyAsnAsnThrThrGlnSerThr----- 880
QY 3586 GATACAGAAATTTTGATGATGTCACTCAACAAATGATTTGTGATGAAAAAGTAT 3645
DB 880 ----- 880
QY 3646 GGTAAAGATACGAAAAATATGCTTTAGAGATTAACACAGACCATGATGCTGCTGT 3705
DB 881 -----ThyLysPro-----GlyAlaAla 886
QY 3706 GGTGTAAAGTGGATGGAACGACAAAGGTCACAGATTAACAAACGAAAAACGGCA 3765
DB 887 GlyThrProSerGly----- 891
QY 3766 GAAAGCATACGGAATGTAACAGTCGATGATATCTTAAGAAACGATGGAAGAA 3825
DB 892 ---LysAspThr----- 894
QY 3826 CAAGTACAGATTTGTCTCCAAAAAGATATGATGATCCCGATTGGCAATGCGCA 3885
DB 895 -----Gly 895
QY 3886 AATATTAATTTAGTGAAGACCTGCTGTGTATGCCCTAGAACACAAAGTTATGTC 3945
DB 896 SerIle-----CysValProProAlaGatGatGlyLeuLys 907
QY 3946 GTACATTTCTTG----- 3957
DB 908 ValGlyLysLeuHisAspTrpAlaGlyGlyLysThrGlnAlaLysSerGlnGlyThr 927
QY 3958 -----GCAATGATTAAT----- 3969
DB 928 SerGlyGlyLysThrProSerGlyAsnGlnLysSerProSerGlyLysLeuProGln 947
QY 3970 -----GAAATTAATAAATTCACATGTAATTAATAAAGAGCTTATC 4017
DB 948 GlyProThrProGlnThrLysGlyThrProGlnSerSerLeuHisAlaPheVal 967
QY 4018 AATCTGCACACAGCAAAATCTCTCATGTATATTAATAAAGT----- 4068
DB 968 SerProProAlaGlyLeuArgPheLeuProThrHisLysPheLysGlnGlnTrpLysAla 987
QY 4069 -----GATGTTGAAGCA 4080
DB 988 GlnHisGlyAlaGlyAlaThrGlyGlnGlnThrIleLeuLysThrLeuAspGlyGly 1007
QY 4081 AATGAATCTC-----GATTAAGATTAATAAAGAGCAAAATCTCTCCGATTTTGAATCC 4137
DB 1008 GlnGlnThrProAspLysLeuLeuLysThrGlyHisIleProProAspPheLeuAlaGln 1027
QY 4138 ATGTTTACACATTTGAGATATAGAGATTTTATTTTGA-----ACAGATATA----- 4188
DB 1028 MetPheLysThrLeuGlyAspTyrArgAspIleLeuValGlyAsnThrAspIleVal 1047

QY 4189 ----TCAAAAGCTCATGCTGAGGCAATTAACCTAAACAGCAAAATAGTTCT----- 4236
DB 1048 HisThrSerGlyAsnLysGluAspMetGlnIleMetGlnAlaIleGlnLysLys 1067
QY 4237 -----CTTTCAAAAATGATGACCAAAAATCTCTTAATGGAAGAAACA----- 4278
DB 1068 GlnIleLeuProThrSerGlySerSerProSerProAlaGlyAlaThrGlnThrHis 1087
QY 4279 -----CGCCAAAGATGCTGACAGACAAATGTCATGCAATATGGAAGCT 4323
DB 1088 SerValGlnAsnProAlaGlyLysThrTrpAsnGlnAsnGlyLysIleTrpGlnGly 1107
QY 4324 ATGCTATGTCACACTGTA----- 4341
DB 1108 MetValCysAlaLeuLeuThrTyrAsnThrAspThrProSerGlyThrAlaProThrGlnIle 1127
QY 4342 ----AAATTTGGCGCAAAA-----AAAGATGATTTTCGCAAAAC-----TAC 4380
DB 1128 GlnGlnValAlaThrThrLysLeuArgAspGlnAsnSerLysAsnProLysIleProGlnTyr 1147
QY 4381 GGTTCACAAACAGCTCAAAATTTAGTGAC----- 4407
DB 1148 LysTyrAspGlnValLysLeuAspAspThrSerAspAlaLysThrThrGlySerProVal 1167
QY 4408 -----AAAGCCACACTTTGGAGAGAAATTTGCCAAACGACCCAGTTTACGA 4455
DB 1168 ProSerGlyLysLysIleThrProLeuThrAspPheIleSerArgProProTyrPheArg 1187
QY 4456 TGCGTAAACGAAATGTCAGACGACTATGCTATACAGCAAAATTTGGAAGAGTGT 4515
DB 1188 TyrLeuGlnGlyLysThrGlyLysThrPheCysLysGlnArgLysAspGlyLysIle 1207
QY 4516 CAGCAAAAATGTAACTGA----- 4533
DB 1208 LysGlnGlyLysArgGlyAspArgThrGlyHisGlnHisCysSerGlyAspGlyTyrAsp 1227
QY 4534 -----AATGACCA-----TTGAAGTGTGATACGAA 4560
DB 1228 CysThrArgThrAspAlaAspArgAsnAspLysPheValAspLeuAsnAsnAsnArgAsp 1246
QY 4561 TGTAATAAGAAATGCGAGACTACGTAAATATATGAAA-----AAAAAAAAGATGATG 4617
DB 1247 CysHisIleGlnCysAspGlyTyrArgLysTrpLysPheLysPheAspGlyLysHis 1266
QY 4618 CCACAAATTAATTAATTAACAGATGAACGCGCAAAAAGATTCATGACAAACAT 4677
DB 1267 LysGlnGlnLysLysTyrGlnGlyLysTyrAspLysLeuThrLysAspLys----- 1283
QY 4678 GGTGTAATGCTTACAGCTACTGGAACGAAATGCAACAGATTACTTGAACGAGAAATTT 4737
DB 1283 ----- 1283
QY 4738 ACTGCTAGTGTGCTGATTAAGCTGGAAGTGCCTCTGTGTAACAAAGAAATATCATTTG 4797
DB 1284 ---SerSerGlyLysAsp----- 1288
QY 4798 TTGAAAAAAGAGCTTACTATGATCCGACAAACATTTGTGGTGCACAAAATTTATTTGAA 4857
DB 1289 -----AsnAsnCysCysLysAspIleGlnLysHisLysSerAlaAlaValPheLeuLys 1306
QY 4858 AATGACGACAAATATACTAATCATTTGCAAGTAAGATAGTCAAAAGATTAAGTAAGAG 4917
DB 1307 Gln-----LeuLysHis 1310
QY 4918 GCAAAACAGGTGCTATTAAGTGCACAAACAAAGTCTCTATATACATTAATTAATCTGAA 4977
DB 1311 CysLysAsnGlyGlnThrSer-----GlnAsnLysGly-----AsnGlnGlnAspGlnLeuAsn 1328
QY 4978 GAATTTGACTGAAAGATGCTTTT-----TCCCTTCCTGCACTGATATATGTTTTCATCC 5033
DB 1329 LysLeuAspPheAspLysIleProGlnThrPheSerProSerThrTyrCysLysAlaCys 1348
QY 5034 ATTGATGCGCAATTAATACAGATCCAGAAAGTTAAAGATGAAGAAATGGTTGCAAAAGATTT 5093

Db	1349	-----	Proval	1350	----	513
OY	5094	GATGGAATGGCGCCACGGAAGGAGTACAAATTTGGTGCATACTACAAAGAAAAAGAA		5153	----	5153
Db	1351	TyrGlyValaIscysAsnGly	-----	AsnLysArg	1360	----
OY	5154	AAAAGAGAAATATAAAAACGTCGGATCGGCACAAATATCTTATGAGAGTCCCGCTTAG		5213	----	5213
Db	1361	GlyArgGlyGlyThrAsn--GlyCysThrThr	-----	AsnAsnGlu-ProGlnAsnLysGly	1370	----
OY	5214	TGCTATGAATATATGTTTTTATGATTTAAGAGATATATCTAGTCTTATATTTGGA		5273	----	5273
Db	1371	-----	AsnAsnGlu-ProGlnAsnLysGly	1378	----	1378
OY	5274	AGATGAAGAAACAAAGACCGAGGAAAAATTTGAAGAAAAATTTAAGAAAAATGGAACATC		5333	----	5333
Db	1378	UAsnAsp	-----		1380	----
OY	5334	AGTTGGCAAGAGATGATAGTACTACAGA	AATCCGGGTAGTCTCGCGG	5384	----	5384
Db	1381	-----LysGlyAlaAlaSerThrIleSerIleLeuIleAsnAspGlySerThr	-----	1396	----	1396
OY	5385	AAAATTTTCTGGACGAAAAATTAAGAAATGTGTGGAACGCAATGATATCGGGTACAA		5444	----	5444
Db	1396	-----		1396	----	1396
OY	5445	ACGTGTAGGATGATGAAATATGGAATATAGTCAAGAAAGTGAATCTAATAAAA		5504	----	5504
Db	1397	-AsnGlyAlaThrAsnGlyThrThreGlyThr	-----	ThrAspGluThrLeuLysGly	1413	----
OY	5505	ARGTGTTCTGACCTTCAGATGATGATTATCTATG	GGGAAAAATCGCGATGA	5558	----	5558
Db	1413	UCys	SerAspLysThrAlaPhePheLysGlyLeuArgLysGlnGly	1428	----	1428
OY	5559	AGGTACTGGGTATCAAGTTTCTTCGATGGTTTCCCAATGGGCTGAAGATTTTGGCAACA		5618	----	5618
Db	1428	UtrpHrCysGlnLys	-----	LysTyrGlyValaAsnGlnCys	1440	----
OY	5619	TAAACAAAAGAAATTTGGAGAAATTTGGTAGGGCTGTATGATTTACTTTGGGGATATA		5678	----	5678
Db	1441	-----	AsnLeuThrAsnArgValaAsnAsp--ThrTyrPheAspLys	1453	----	1453
OY	5679	TGAAGATAAAAAGAAATGTACAGATTCGCGTACACAAATATATAAAATTTATTAGTGA		5738	----	5738
Db	1453	sAsp	-----	IleValaPheAsnGluPheGlnArg	1463	----
OY	5739	GTGGAACCCACAGTATGAAAAACAATTCAAAAATATGTGTGAGATTAAGCAAAATATA		5798	----	5798
Db	1463	gTrpLeuArgTyrPhe	-----	ValHisAspPyrTrsIleLeuLysHisLysIle	1479	----
OY	5799	TTCCGAGCATCCTGTGGCAAAAGATGCAGAGAGCGCTCGCAATATTTAGCAAAACAAT		5858	----	5858
Db	1480	-----	AspProCysIle	1483	----	1483
OY	5859	AAAAAAAATTTGTGAATAATAAGGTGAGATTTGAAATATATAGTATAGTAAAGATGTGTC		5918	----	5918
Db	1483	eLysLysGlyLysGlnAspLysThr	-----	GlnHisLysCysIle	1496	----
OY	5919	CACACAGGATTAATGATGTATTAAGTCAAAATATATGCCGCATCATTAAGCATGAMCC		5978	----	5978
Db	1497	-----	AsnGlyCysAsnIle	1501	----	1501
OY	5979	AAAAAGATTGAAGAAAGTGTATTTGTCTCAAGTCCACAGAGTCCACACAGCTGTAGCAAG		6038	----	6038
Db	1502	-----	LysCysGlyCys	1507	----	1507
OY	6039	GGAACACCGTACACACAGGGTATCACTGATATCAAAAGCGACGCGCATCGAAAAAGAAC		6098	----	6098
Db	1507	-----		1507	----	1507
OY	6099	GAAAAACGCCCGCTACAAATACAGCGGAAAAAGTGGAAAAATCTTAAACAAGAAATGCG		6158	----	6158

[illegible]

[illegible]

QY	8229	TGTTTATTACTCAACGGTTTCGATGNGATAGCTGAATGGTGTGAATATTATGTGAAGC	8228
Db	1938	raspyrlleProbinProphearpyrmetcInglutIserGIuserPhecslytle	1948
QY	8289	ACTGATGCAAAATTTGGAAAAATTTAAAAAATATGATGTGATCACTGn-----AAAACATCGA	8345
Db	1948	uLeuansgluIuKetglutInpheglutInrPheglutInrCysgluGluCysLysLysasnSerIl	1968
QY	8346	CAGATGCAAGATGATTTATATGCAAAATAGTGTGAACAGTGTGAACAGAGATGTCAAGA	8405
Db	1968	ethrcysgluasnpsrpsrpyasnglYrthranCysgluasnCysLysasnGlnCysgluYl	1988
QY	8406	ATATATAAAATTTTGTCTTAAATGCAAAATCTTATTCGATATACATCAAAATTAATACA	8465
Db	1988	styrIysLysLeuIleHisntrPrpLysleuGlyPheasp-----LysTyrLy	2004
QY	8466	AGAATGTATTAACAACCATATAT-----ACAAAAATCTACTCTTATGATCA	8513
Db	2004	sgluIeIyrrasnIu-----IleYrrasnasnLysasrSerLysIleasnSerasnGluYl	2023
QY	8514	TGTTTAAAAATTTTGTATACAAAAAGTTTGAACATTTTAAATGCAATGn-----TCGTGTGAG	8570
Db	2023	rPheLysLysPheleuIuLysleu-----LysasprLysCysLysgluLeuasnSe	2040
QY	8571	CTTTTCGATATATCTTCATGCAAAACAAGTAGTGTGAAATTAATTAATTAAGAAAATGA	8630
Db	2040	rserasprLysCysLysleaspGluIatrrHisCysThrLysYrLysPheSer-----AsnSe	2059
QY	8631	TGCTTCTCCAAATTAAGCAACATATGTCTTTCGAAGAAACCAAAAAAGTTATAAGAAGC	8690
Db	2059	rgluasnLysasnHisasnasnYrralarpheLysasnProPolrLysgluYrGluLysAl	2079
QY	8691	TTGCAGTTGTCACTACTACTCTTCAAGAAATCATTTGGATTAATTCCTACAGCAAA-----	8745
Db	2079	acYsLysCysasnpralarp-----AsprProLeuasnPancysPrpLysasrSerIatrh	2097
QY	8746	AACAAGATGATGTAGAAGATTTACAACTTTTACCTTCGTCTGCAAGATGATTANGA	8804
Db	2097	rTyrIuLysAlacYasnThrLysleuProThrLysleuGluSerLysThrPheas-----	2117
QY	8805	TAAATATCTTATATATTGGAACCATACCTTGTCTTAATAGTTCAGATGATACAAAGC	8864
Db	2117	nasnaspasrpsrserIrrpsrIrrPhevalGlnThrSerProLragsAsnThrGl	2137
QY	8865	TGTATTTGATTCCTCCAGAAGAGACATTTATGACAGCCATACGCATCATTAATTA	8924
Db	2137	yvalIeuvalProProkarIrragIrrglnIecYsLeuLysasnIleThrhr--LysLe	2156
QY	8925	TAGAAAAAGTATATA--GAAATTTTAAAAAAAACCTTACTCTGCTTCAGTCA	8981
Db	2156	uargserIleclIuLysLleasnpsrPheLysalaglIuLeumethrIserAlatYrAsnGl	2176
QY	8982	AGGACAAATTGTTAGCTCAAAAAATTAATCGAAGAAGAGTTGGCTTTGAGCCAAATGA	9041
Db	2176	uGlYrLysleuLeuGysgluLeuYrLysLysasrPaIrrLeuGlnalIamelyl	2196
QY	9042	ATATATGTAATGACAGATATTCCGATTAATTAATTAAGACATGATPANGATGACATCTTCT	9101
Db	2196	stYserIrrheYrrasprlyrrglYasprIlevalYsglyThrAsprleuIleSerThrIalPr	2216
QY	9102	ATCTGAAAAAATTAAAAAATATTT-----GAAACATCAATGAAGC	9143
Db	2216	oleuasnpsrLysleuYsThrLysleuasnValIleuLeuLysglYasprGlyThrAsnGluIl	2236
QY	9144	AACGCAAAATGCTAAAAACATGGTGCGAAAAATTAATAGCTGCAGATATGGCAGCTATGTT	9203
Db	2236	eLysgluasnpsrIrrglYsrrIrrThrIrrgluasnarrIrrarYvalIrrPheIrrsalIamelyl	2256
QY	9204	ATGTGATATATAAATFTGCTACTTCACAAAGTAACATTAGATGAAGAAGATGGTGTCAATTCG	9263
Db	2256	uCyselYrlyrrLysalalacIyglYslyle-----GluIrruIrrgaspCysSerIrrLe	2274
QY	9264	AAAGATGATGAAACCTATCATGTTCTTCCTGTGTGTAATGTAATGGCGAAACAGATCAT	9323


```

Db 2274 0-AspaspasnthrhsglnpheleuargtrpheardglutrpserglnhisphcY 2293
QY 9324 TAAGAAAGAAACATGTAAGTATCATTAATAAACAATAATGCTGCTCAACAGAGA 9383
Db 2293 salblyserglinsylleupheasngluvalylsarglucys-----AlaSerAlaG1 2311
QY 9384 TAATTTGAACCGTCAGAAATTATTAAGACAACCTGATGTCAG-----AATGATAT 9434
Db 2311 ncysileileglutryrclgthrileasrprovalcysgluclualacysrthrlnY 2331
QY 9435 TAGAAATATATATGCTGAATATGATTAATAAATACATGGAATCAATATATATAA 9494
Db 2331 rargasptrylilerthrargylsilegln-----G1 2341
QY 9495 ATATAGCAATATCAAAAGATCAATCTTCAGTAAATATAGCAATTAACCATCTGAA--GA 9551
Db 2341 utyrargleuleasnthryrclntrhrasnphbasnclulyslsalaglualtm 2361
QY 9552 AAATGTCACATCATATATAAATCAAAAGATTCATCAATGCGCTTGGAGTTAATGATAT 9611
Db 2361 rlysalaprogltutyrpheasnasplys-----Cysasnasplicysasnscysle 2378
QY 9612 AAATGAAATAGTTACAGAGCAAAATAATATGAATAATGATTCAAAGACTACTAA 9671
Db 2378 userysTyrIle-----AspIleglulyslystrly 2389
QY 9672 AAAATATATCTCGTTTATATTTGTTGAGATGAACACACAAAATCATCTACTAGA 9731
Db 2389 sasmetlyrAspserPhe-----AspaspasnaspIleuylasnlscysilecy 2406
QY 9732 TGGAATATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9791
Db 2406 salggnlllelpsprolysrgrPro-----ProlyslysalgluylsProglucl 2423
QY 9792 CCATAGATATCTTCTTATCAAGACCTTATCTCAACATCAAGAGAGAGAGAGAGAG 9851
Db 2423 uhlstrprosergluclnasphrproproleuproprolysrProaspaspIeudr 2443
QY 9852 TCCT-----AAAAATCATATATGAAAGTAGTATCTGTTGTTAT 9893
Db 2443 oProProAlagluluproPheasnargaspIleuclulysThrIleProPhegllyl 2463
QY 9894 TGTATCGCGGTAGTATGATAGCGCTTCATTCATGAGAAATAATCAATCGCTCT 9953
Db 2463 ealaleuAlaleuaglserIlealalpheuleuclulyslysrThrLysSerSerVa 2483
QY 9954 G---GACTTGTTCGCTATACGAATATCCGCAAGAGAGATGGAATGCTACGTGGA 10010
Db 2483 lgllysnleuphegnllleleuhsilrProlyserAsptryrAspIleProThrLysle 2503
QY 10011 ATCCAAAAATAGTATCATCTCAATATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 10070
Db 2503 uSerProasnargylrIleProtyrThrserIlysrlysrargylsargylrIlely 2523
QY 10071 TATGAGAGAGATACT-----AGTGCAGATGAGATTAATATGTCGAGACTTATC 10121
Db 2523 rleuclulglaspserglYThrAspserGlyrThrAspHisIlyr----- 2538
QY 10122 TTCCCTGATATACT---TCATCCGAAAGTATGATGAGAAATGATTAATGATAT 10178
Db 2539 ---SeraspIlethrSerSerSerSerIusserclutryrclglululeuaspIleasnaspI 2557
QY 10179 ATATATACAGTATGCTTAATATTAATAAATGATTAATAAGTATGATGAGAGAGAGAGAG 10238
Db 2557 eTyraIleProcllyserProlysrlysrThrleucllelgluValleuclulProserG1 2577
QY 10239 AAGGATATACCAAGTGAT-----GATACACCAAGT-----AA 10271
Db 2577 yAsnasnthrThrAlaserglYlysasnThrProserAspThrclnasnaspIleglins 2597
QY 10272 TGATACACACAGTACGATTAATTTATGATGAGATGAGATGAGAGAGAGAGAGAGAGAT 10331
Db 10331

```

```

Db 2597 naspglylleProSerSerlysrIlethrspasnglutrpasnThrleuLysaspIuPh 2617
QY 10332 TGATATCATATATTA-----CCAAATACGAGAACCAAAATTAATATACAAAAGTGC 10382
Db 2617 eileSerAsnMetleucllnasncluproasnThrIglPro-----AsnMetleuclly 2635
QY 10383 AGATATCCCAATGATACAGAACCTTAATCTTATATTTCTGATATCCGAAAGAAAACC 10442
Db 2635 rAsnValaspasnAsnthrhIsProThrThr---SerArgHisAsnValaglululysPr 2654
QY 10443 TTTTATATATCATATCATGATAGGATTAATATCTGGAAGAAATATGATTAATAT 10502
Db 2654 oPheileMetSerIleHisAspArgspIeudrYserglYglulutryrSerlyrAsnVa 2674
QY 10503 TATATGAGTACTAATATCAATATATGATATTCATGAAATGCTAGAAATGATTCATTA 10562
Db 2674 lAsnMet-----ValasnasnaspIleProIleSerAlaArgasnclYAsnlyse 2691
QY 10563 AGTATAGATTTAATTAATGATTCATGCTGCTAAACCATATGATATATATGATGA 10622
Db 2691 rclYlleaspIleucllnasnaspSerleu--AsnSerAsnlysalaspIletryrAspIu 2710
QY 10623 GTATG 10628
Db 2711 leuLeu 2712

```

RESULT 6

128161

Hypothetical protein FC3-variant1-1 - malaria parasite (Plasmodium falciparum) (fragme

C/Species: Plasmodium falciparum

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C/Accession: T28161

R.Hernandez-Rivas, R.; Mottel, D.; Sterkers, Y.; Peterson, D.S.; Wellens, T.E.; Scher

Mol. Cell. Biol. 17, 604-611, 1997

A>Title: Expressed var genes are found in Plasmodium falciparum subtelomeric regions.

A/Reference number: Z20483; MUID:97154495; PMID:9001213

A/Accession: T28161

A>Status: preliminary; translated from GR/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-2647 <HR>

A/Cross-references: EMBL:U06795; NID:91794255; PID:91809295; PIDN:AAC47438.1

A/Experimental source: strain FCQ27/PNG

A/Genetics:

A/Introns: 2158/3

A/Note: FC3-variant1-1

Alignment Scores:

Pred. No.:	9.18e-127	Length:	2647
Score:	2727.00	Matches:	922
Percent Similarity:	34.82%	Conservative:	399
Best Local Similarity:	24.30%	Mismatches:	810
Query Match:	14.02%	Indels:	1664
DB:	2	Gaps:	129

US-10-087-013-1 (1-10628) x T28161 (1-2647)

```

QY 91 AGTCACAAAAGTGCAGAAAATTTTGAAGCTATGTCGCAAAATATAGACATCCATCA 150
Db 9 ThrasnlysserAlalysgluValleuaspIucllelglulthrIleglin----- 25
QY 151 AAATATGCAAAAGCAACATGTGGATTCG-----TTGAAGGGAGATTG 192
Db 26 -----LyslysalahIssetAspAlaAspThrPheargsergluleuylsGlyAsnph 43
QY 193 ACGAAGACAGAAATTCGTGGTGGT-----CCTTCAGCCAGTAATTAAG 237
Db 44 GlyclulalysrheYrAsnglyglulilemetclnProasnserlysrleu----- 61
QY 238 CATTAATATTATATCATATCATGATTAATTTAGATCATAGAGACATATAATTAACG 297
Db 62 -----CysgluleuAspHisThrIleasprhrAsnValThr 73
QY 298 TATGATGATGGAATTTGAGACATCCTTGCCATGTAGAGAACAAACGATTGATGAA 357

```



```

OY 2395 AGAAGAGCTTCAAGCAATTATGATGATTAATGATTAACATCTGTAATCTT 2454
Db 795 ThrIleuLeuLeuAsp---IleCysSerIleThrAspHisSerAsnAlaLysArg 813
OY 2455 GGTTCCTCAATGACCATGTGATGAGCAAGGCGACAGGTGATGATATCAACAAGATT 2514
Db 814 GlyHisThrAspGlnProCysLysArgLysAspSerLysValAsnValLysAsnArg--- 832
OY 2515 GTCTAGCAATGCAATGGAAGTGGATCCGACACATGGGTGAAGTCGCAAGATGTT 2574
Db 833 -----ArgTrpMetAspThrAlaGlyPheIleSerAsnThrTyrLysAspIle 848
OY 2575 ATTATGCTCTTACAGACACATATATGATACATTCATTTGAAACATTTCAACAGAT 2634
Db 849 TyrMetProProAlaArgGlnHisPheCysThrSerAsnLeuGlnTyrLeuGlnThr 868
OY 2635 GATCACCACCTTAATGATATGTTGAT-----GATTATGTTAATTAATCTCTT 2685
Db 869 AsnLysLeuLeuAsnGlnLysAsnAspIleAsnGlyAsnProAsnIleIleAsnAspSerPhe 888
OY 2686 TTGGGGGATGTTCTTCTATCAGCAAAATATGAGCAAAAGATTAATATGATGATTA 2745
Db 889 LeuGlyAspValLeuPheAlaIleAsnTyrClnAlaAspPheIleLysLysMetTyrAsn 908
OY 2746 GAAAAAGATTAACCTAAAGGCCCCCAAGAACTACTGACCCAAACACACAGACACTATC 2805
Db 909 LysGlnAsn-----AspTyrLysAspAsnAlaThrIle 919
OY 2806 TTCTGAGCTATATGATTTTTCAGATATATGATGATATATATGAGGAGAGA----- 2859
Db 920 CysArgAlaMetLysTyrSerPheAlaAspLeuGlyAspIleIleGlnArgGlnHisIle 939
OY 2860 -----GATCTGTGGGAAAGAAACGGTGCATG 2886
Db 940 CysArgIleMetIleValGlnArgValLysHisGlnIleSerGlnArgAsnPheLeuIle 959
OY 2887 GTAAGCTCAAGACATTTGGAACACTGTTTGTGTAATATCATTAAGTCATCAAGGC 2946
Db 960 LeuSerLysLysAsnIleLeu-----AlaPheLysGlnIle----- 971
OY 2947 AAAGCAATGATTAATATATGATGATGCCCCCAATATTTAAATTTGAGGAAATTTG 3006
Db 972 -----TyrLysGlnAspThrPro---TyrThrLysLeuArgLysAspTrp 985
OY 3007 TGGGAAGCTAATAGAGCCAAAGTATGCGGAGCCATGAATGTGATTAATAATATTGAAG 3066
Db 986 TrpGlnAlaAsnArgLysLysIleTrpGlnAlaMetGlnCys----- 999
OY 3067 GATTAATCGGACACCAATCAACACAAAGTATGTTGCGATATAGTATGATACACCA 3126
Db 1000 -----ProThrProAsnGlnYserPheProCysLysSerTyrHisIleGly 1014
OY 3127 TTGATGATTAATATCCACAAAATTAAGATGATGATCCGAAATGGGCAATGTACTGC 3186
Db 1015 LeuAspAspTyrIleTrpGlnArgLeuArgTyrTrpMetThrGlnTyrAlaGlnIleTrpPheCys 1034
OY 3187 AAGTGCAGAAAAGAGATGATAGTTGAAGAGAGAGTGAAGATGTGAAGATGAAG 3246
Db 1035 LysGlnGlnLysGlnTyrClnLysLeuValSerAlaSerAsnGlnCysLysAsp--- 1053
OY 3247 GATTAATGCTCAAGCTGTACAGAAAGAGAGTGTACAGGTGTGACAAAGTGCACAG----- 3301
Db 1054 GlnArg-ValLysValAlaArgIleArgValHisAsnValGlnArgAlaCysLysHisLeu 1073
OY 3302 -AAGCTGTATGATTAATATGATATATAGCATTTAGGAAAGCAATGGAATTAATA 3360
Db 1073 LysIleIleLysAsnLeuLeu---IleHisGly-----LysGlnGlnTyrAspLysMet 1090
OY 3361 TCAGATAATCAAGAATATCATGACACACACAAATGTCTGTAGAAATAGGATAT 3420
Db 1091 GlnIleLysTyrLysLeuLeuTyrLeuGlnAlaGlnThrThrAlaAlaAsnGlyGlyPro 1110
OY 3421 GAAGCTTCAGTACTGCCAAAAATCATATAGACAGAAATGTAATTTTGTCCGAA 3480
Db 1111 AspThrTyrSerLysLeuValAspGlnAsnGlnLysProValValAsnPheLeuPheGln 1130
OY 3481 TTATACCAACAAAATGTTGGCAAA-----AGTAATAAA 3513
Db 1131 LeuTyrLysGlnAsnGlnLysIleGlyAsnProAlaGlyAspThrProArgAlaLysArg 1150
OY 3514 AGTGTACTAGTGAAGAAAGTGTGATGCTGACTACTACACCAAGTGAAGAAATGTTGA 3573
Db 1151 SerLysArgGlnThrAlaProAlaSerValAlaLysAsnAspValLysSerThrAlaIle 1170
OY 3574 GCATATCTGCAT---GATACAGAAATTTTGAATGTGATGTGCACTCAAAATGATTTGT 3630
Db 1171 GlyTyrValHisGlnIleMetGlnLysProHisMetGlnCysLysThrGlnThrClnPheCys 1190
OY 3631 GATGAAAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3690
Db 1191 ---GlnLysThrAspLeuGlnIleTyrAsnGlnLysAsnTyrThrPheLysAsnProProGln 1209
OY 3691 CATGATGCTGCTGTGTTGTAAGAGTGAATGCAACGACAGAGGTACAGATTAATAAG 3750
Db 1210 TyrLysAspAlaCysIleCysAsnThrArgProProLysGlnLysAspSerArgLysArg 1229
OY 3751 AAAAAA-----AAAGCGAAGAAAGATRCG 3777
Db 1230 SerGlnAspSerAspLeuGlnLysValLysGlnThrLysValGlnLysAlaThr 1249
OY 3778 GAA----- 3780
Db 1250 GlnAspAlaValAspThrGlyProProProAlaProLysGlnAlaThrThrLeuAsp 1269
OY 3781 ---TGTAACAGATGATATATACTTAAGAAAACGTGAAAGAAACAGTGAAGAT 3837
Db 1270 ValCysProIleValAlaGlyValLeuThr-----LysGlnAsnLeuGlnAsn 1285
OY 3838 TGTCAATCCAAAAAAGAAATGATGATGATGCC---GATTGCAATGC----- 3882
Db 1286 AlaCysProThrLysTyrGlyProLysAlaLeuProThrSerTrpLysIleProThrGln 1305
OY 3883 -----GGAATATTAATTTAGTGAAGAC 3906
Db 1306 LysThrAsnAlaIleThrGlySerGlnLysSerSerGlyAsnGlnAlaLeuGlnArgAla 1325
OY 3907 CCTGCT-----GTCGTGATG 3921
Db 1326 LysArgAlaThrValGlnSerGlySerProValThrSerAsnSerGlySerIleCysIle 1345
OY 3922 CCCCCTAGAACAAAAGTTATGCGTA-----CATTTCTTGCAAAATGATTAATGAA 3972
Db 1346 ProProArgArgArgArgLeuTyrIleGlnLysLeuHisAspTrpAlaSerGlyAsnThr 1365
OY 3973 ATAAAAAATTAACAATCAAGTT----- 3996
Db 1366 ValValSerGlnAlaGlnIleThrProGlnLysGlyThrSerSerProSerGlyLysGln 1385
OY 3997 -----AATTTAAAGAGCTTTCATCAAAATCTGACAGACGCAAAATCTTCT 4044
Db 1386 ThrProSerAspLysLeuArgThrAlaPheIleGlnSerAlaIleGlnIleThrPhePhe 1405
OY 4045 TCATGATTAATTAATAAATTAAGATGATGATGATGATGATGATGATGATGATGATGAT 4104
Db 1406 LeuTrpAspArgTyrLys-----LysGlyLysAlaIleAlaLysLysGlnLys 1421
OY 4105 GAAGGCAAA----- 4113
Db 1422 LysLysGlnMetValAspTyrSerProLeuSerThrAlaAspProHisAsnAsnProVal 1441
OY 4114 -----ATCTCCCGCATTT 4128
Db 1442 SerLeuValIleAlaProAsnProAsnTyrAsnLysThrCysValIleProProPhe 1461
OY 4129 TTGAGATCCATGTTCTTACACATTTGAGATTTTAAATTTGAACAGAT--- 4185

```



```
QY 6301 AATCAAAAACGTATGGACAAATATCTAAATGGGGTGTATGTAGTAGTAAAGAA 6360
Db 1737 ----- 1737
QY 6361 AATGAAATGSCATATGTATGCCCTCTAGAGAAAAAATTATGTATTAATATACAA 6420
Db 1738 ---LysasnGlyProCys----- 1742
QY 6421 TATTTAAATTAATGAACTGAAATTAACGCTGACATATGATTAAGAGCTTTTATTA 6480
Db 1743 -----LysasnAspAsn----- 1746
QY 6481 TGTGCACCAATAGAAACTCAATTTTGTGTAAATATATTAATGAAATCCTGCACA 6540
Db 1746 ----- 1746
QY 6541 GAAATGATTCGAAATGAGACAAATTCAGATTAATTAAGAAATATGTATATACA 6600
Db 1746 ----- 1746
QY 6601 TATGTGATTAATAAGATATGTTTTTGAACGTATATTTCTAATGATAAAAAATTATA 6660
Db 1747 -----ValAspAspSerCysGlyLysAsn----- 1754
QY 6661 ACTGTAAACAATATGTATACACACATTCATGAAATATATAGAAAAACAGATAAA 6720
Db 1754 ----- 1754
QY 6721 AAAAAAGATGAAAGATTAAGTAAATATTTTGGAGAAAAATTAATTTATTTGGAA 6780
Db 1755 -----LysIlePhe----- 1757
QY 6781 GGAATGATATATGATTAATTAATATCTATCATCTCAGACAGAAAGAAAAAATTAGA 6840
Db 1758 -----AspGluAsn-----Gly 1761
QY 6841 GATTAATACAGTACATGACATGACAAACTGACGCTTCCTCGTAAGATTGTATAA 6900
Db 1762 ---AspThrPheLysTyrThr----- 1767
QY 6901 AGGCCCCAATTTTGTGATGTTCCACAGAAATGGCAGAAAGAAATTTGTAATAGAGAG 6960
Db 1768 -----GlnTyrCysGly----- 1771
QY 6961 GAACAGTTGTTAAATTTGAGCGGGCTGTAGAAATATAGATGTAATGTATGATGAC 7020
Db 1772 -----ThyCysSerLeuAsnGlyPheLysCysAsnGlyAspAsp--- 1784
QY 7021 GGTAAAGACAAAGATGTGCAGAGCGGTGTATACATATCAAAATTTTATTAAAGAGTG 7080
Db 1785 -----CysArgValArgThrAsn----- 1790
QY 7081 AAACTGATATGAAAGACAAAGAGAAAGTTCAAAAAGATTAAGATGCGCAAAAGTAT 7140
Db 1790 ----- 1790
QY 7141 AAGGATATTCCTTCTACTGAAAGACATATAGAGAGCAACATGTCATGATATTTA 7200
Db 1791 -----ValThrCys----- 1793
QY 7201 AACTGAAATTAAGAAATTAATGTGCAATAAGATTTCTTGTATGCAAAAACCTTCT 7260
Db 1793 ----- 1793
QY 7261 TCACAAATACCAAAAACAAACAAATCACAATCATCCGATGCTAATGATATGCCGAA 7320
Db 1794 -----AsnGlySerAsnArgThrThrThrIleThrAlaAspAspIle----- 1807
QY 7321 TCGGTGATTAATGTTCCAGAAAGATTTAAACAAGTGTAGTGTCTGAACTTCAAAAAG 7380
Db 1808 -----LysAsn 1809
QY 7381 GGATCTATGATTCATACAAAAAATTTACTGAACTTAAATACCTATGATTTGTAGAG 7440
Db 1810 Gly----- 1810
QY 7441 AAAGCAGATATTAATTAATCTAAAGACAGACAGAAATATATG-----GATATT 7488
Db 1811 -----GlyAsnSerAlaGluIleAsnMetLeuValSerAspAspIle 1824
QY 7489 ACCTTAAGAAAAATTTTATACCTATGAGTCTACAAAGAAAAAGAAAGTAAAAATAGT 7548
Db 1825 AsnSerGlyAsnGlyPheAsnAspLeuGluAla----- 1835
QY 7549 TGGACTAATATATATTCCTTGCATCCTTAAGAAACCTTATGCACTGATTAATATAGA 7608
Db 1835 ----- 1835
QY 7609 AGAAGAAACCTTGTGAAATATAGAGAGAAATCCTTTAAG-----GTAGATTATGAA 7662
Db 1836 -----CysLysAsn-----AlaAsnIlePheLysGlyIleLysGluAsnLys 1849
QY 7663 TGGAAATGTTACAAAAATTTCAAGTCTATCATCAGGAAAAAAGATATGTACCTCCA 7722
Db 1850 TrpLysCys-----ValTyrPheCysLysSerAspValCys----- 1861
QY 7723 AGAAGAAACATATGTGCTTAAGGAATTTAGATGAATTAATTAATGAAAGCTTAAGAT 7782
Db 1861 ----- 1861
QY 7783 AGTATTAATTCCTTAAAAATGCTTGTGCGAACCTGCACGAAATGAAGAAATAGACATA 7842
Db 1861 ----- 1861
QY 7843 AAAAACTTCAACTCAGAGAAAGGGTGGCAATGATCCAAATATGATATGATTAATAT 7902
Db 1861 ----- 1861
QY 7903 AGTTTCGTGATCTGGGTGACATAGTTAGAGAAACAGATATGTTACGAATTTGTGTAC 7962
Db 1861 ----- 1861
QY 7963 TTACCTCCGTAGAAATTAATTAATTAAGTTTTGAATACATATATGAAAAATGAGA 8022
Db 1861 ----- 1861
QY 8023 AATAAATAAAGGTAGAAATTAATACACAGATGACAAAGCTTTCGTCTGTTGTGG 8082
Db 1862 -----GlyLeuLysLysAsnAsnAspIle----- 1869
QY 8083 GATGCTAATAGAAAAAGATATTTGGAAGCAATGACGTGCAGAAAGCAGACAGATGCAAAA 8142
Db 1870 AspGluAsnGlnIleIleLeuIleArgAlaLeu----- 1880
QY 8143 CTTTGTGAAGAAAGAAATGATGATTTGAACCGCATTAACATTAATACAAAGATAGT 8202
Db 1881 -----PheLysArg----- 1883
QY 8203 GCACATAGAGACATCCACTGTTGATGATTAATATACCTCAACGTTTCGATGATGACT 8262
Db 1883 ----- 1883
QY 8263 GAATGCTGATTAATTAATTTTAAAGCAGCTGATGGAAGAAATTTAAAAAATCA 8322
Db 1884 ---TrpLeuGlnTyrPhe-----LeuAspAspIlyrAsnLysIleArgLysLys 1898
QY 8323 TGTGATCACTTAAMACATCTGACAGATGCAAGAAATGATATGATGAATAATAGTGA 8382
Db 1899 LeuAsnProCys-----IleAsnAsnGlyGluLysAla 1909
QY 8383 CAGTGTAAAAAGATGTCAGAAATTAATAAATTTGTTCTTAATGAAATCTCTATTC 8442
Db 1910 IleCysThrAsnGlyCysValGlu-----GlnThr----- 1919
QY 8443 GATATACAAATCAAAATTAATTAACAAAGATTTGTATGAAACAACATATATCAAAAATCTCT 8502
```


OY	10486	GAAATAGCTTAATAATTAATTGAGTACTAATCTAATAATGAATTCCAATGC	10545
Db	2368	GUUYYrasVasValasmet---ValasnthrMetaspilreproileanser	2386
OY	10546	AGAATGATCTTANAGGGTTACATTTAANTATGATTCCT-AGRGIGCTTAACCT	10604
			:::
Db	2387	HlsanasnValtySercllyleaspeulleasnaspthrLeuserclylasmcIuhIs	2406
OY	10605	ATTGATATATATGATGAAGTATTG	10628
Db	2407	IleaspietYraspcIuneuou	2414

RESULT 7

variant-specific surface protein 2 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence,revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: J28626
R:Su, X.Z.; Heattwale, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson, D.S.
J. Mol. Biol. 182, 89-100, 1995
Title: The large diverse gene family var encodes proteins involved in cytoadherence and
Reference number: J20487; MUID:95330813; PMID:7606788
A:Accession: J28626
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2664 <SU>
A:Cross-references: EMBL:J20487; NID:g886376; PID:g886378; PIDN:AAA75398.1
C:Genetics:
A:Introns: 2197/3
A:Note: var-2

Alignment Scores:

Pred. No.:	9,71e-127	length:	266
Score:	2726.50	Matches:	906
Percent Similarity:	34.34%	Conservative:	393
Best Local Similarity:	23.95%	Mismatches:	879
Query Match:	14.01%	Indels:	160
DB:	2	Gaps:	122

US-10-087-013-1 (1-10628) x T28626 (1-2664)

QY	97	AAAGAGCAAGAAATGTTTGGAACTGATGCCAAAATATGAACATCATCAAAATAT	156
		: : : :	
Db	17	AtgSerAlaIlyshIsLeuLeuAspSerIleGlyLyIsVal	---
		: : : :	---
		-----Tyr	31
QY	157	GCAAAAGAACATGTGGATTCG-----TTGAAGGGGATTTGACG	195
		: : : :	: : :
Db	32	AspLyIsValIhIsGlyAspAlaLeuGlnProSerAsnGlyLyLeuGlyscLyIhIeuser	51
		: : : :	: : :
QY	196	AAAGAGAAATTTGCGCGTGCCTTCACCCGCAATAAAGCATTAATTAATTATTC	255
Db	52	LeuAlaIlePheGlyLysAlaProGlnGlyLyGlnIhSerGluAsp-----	67
QY	256	TATGCATTAATTAGATCAATAGACATCACTAATTTACG-----TATGATGATGTG	309
		: : : :	
Db	68	---ProCysAspLeuAsnHIsGlyIyIhIstIhThValIhSerGlyTyAspLyIsGlu	86
		: : : :	
QY	310	AATTGACACATCCTGCCATGCGTAAGAGAACAAACCGATTTGATGAAGATGAACATCT	369
Db	87	Asn-----ProCysLyIsAspAlaGProGlnValAlaPheSerTyIhThGluGlyAla	1030
QY	370	GAATGTGA--AATTAATAACGTAATTAATAAAGAAAAAATGATGCGATACCGCTGGCG	426
Db	104	GlnCysAspLyIsSerLyIleAlaGlySerAsnSerAsnLyIsAspGly--AlaCysAla	1222
QY	427	CCACCTGACAGACGACATATGTGTGATTAACCTGGAACTGTCA--AATGATTAAT	4833
Db	123	ProPheAlaGlyLeuHIsLeuCysAspGlnIhIstLeuGlnIhIstIeLyIsAspLyIsIle	142
QY	484	ACCCAAATATTCATGTGATTAATGGAAATGTCACATGATTCACGCAAAATACGAAGGTGAA	5433
Db	143	ThraG-----HIsAsnLeuLeuAlaAspValCysGlyAlaAlaIlyPheGlnIlyAlaGlu	160

[illegible]

Db 1167 ArgThrArgGluCysLeuGlyGlnAsnValIlePheCysAsnAsnGly----- 1182
QY 3655 AAGCAAAATATACCTTTAGAGATAAACACAGACATGATGGTGGTGTAA 3714
Db 1183 AsnAsnGluTyrAlaPheSerLeuThrProHisGluTyrIleHisAlaCysIleCys--- 1201
QY 3715 AGTGATGCAAAACGACAAAGGTTACAGATTAATAAAGCAAAAAAAGCGCAAAAAAGAT 3774
Db 1202 -----AsnGluAsnLysAlaSerSer----- 1208
QY 3775 ACGAATGTAAACAGATGATATACTTAAGAAACGATGGAAACAAAGTACAA 3834
Db 1209 -----ProGluIleuGlyArgSerSerPhe 1218
QY 3835 GATTGTCTCCAAAAAGAAATAGTAATGATATCCGATTGGCAATGCGAAATATAAAT 3894
Db 1219 AspSerHis----- 1221
QY 3895 TTAGTGAAGACCCGCTGTGTATGCCCCCTAGAAACAAAGTTATGCGTACATTTC 3954
Db 1222 -----GlnThrProArgProGluIleAspGluValHis--- 1232
QY 3955 TTGCAAAATGATTAATAAATAAATAATTCACATCACAGTTAATTAAGAAAGCTTTC 4014
Db 1232 ----- 1232
QY 4015 ATCAATCTGACAGACAGAAACATCTCTCATGTATTATTAAAGTAAGATGT 4074
Db 1233 -----SerSerGluGluGly 1237
QY 4075 GAAGAAATGAACTCGATTAAGAATTAAGAAGCAAAATTCOTCCCGATTTTGGAGA 4134
Db 1238 GlnGluIleAspGluSerGluAspGluGluIleSGLu----- 1248
QY 4135 TCCATGTTTCACATTTTGAAGATTAATAGATTTTATTGGAACAGATATATCANA 4194
Db 1248 ----- 1248
QY 4195 GGTATGTGTAGGAGTAATACTAAAGACAAATGATCTCTTTCAAAAATGGTGAC 4254
Db 1249 -----GlnGluValGluGluValHisAspGluAlaAsp 1259
QY 4255 CAAAAATCTCTTAATGAGAAAAACACGCCAAGATGGTGACAGACATAGTCATGATA 4314
Db 1260 GlnLysAla----- 1262
QY 4315 TGGAACTATGCTATGTGACTAGTAATAATGGGGCAAAAAAGATATTACCGAA 4374
Db 1262 ----- 1262
QY 4375 AACTACGTTACACAAACGTCNAATTTAGTACAAAAGACACACTTTGAGGAAATTTGCC 4434
Db 1262 ----- 1262
QY 4435 AAACGACCCCAATTTTACGATGGCTAACCGAATGTACGACGACTATTGTATACAGA 4494
Db 1262 ----- 1262
QY 4495 CAAAAATATTGAGAGATGTGAGAAAAAATGTAATGACCAATTTGAAGTGTAT 4554
Db 1262 ----- 1262
QY 4555 ACAGATGTAAATAGAAATGCGAGACTACGTTAAATATATGAAAAAAGAGGTGG 4614
Db 1262 ----- 1262
QY 4615 ATTCCACAAGATTAATATTACAGAGATGAGCGCAAAAAAAGATTGATAGACACAC 4674
Db 1262 ----- 1262
QY 4675 ATTGTTAATGTTACAGACTATACGAAAGATGACACAGATTACTTGAACGAGAA 4734
Db 1262 ----- 1262

QY 4735 TTTACTCTAGTTGTGGTGTAAAGCTGGAAGTGCCTCTGTGTACAAAGAAATATACAA 4794
Db 1262 ----- 1262
QY 4795 TTGTTAAAAACAGGTTTCTATGATCCGACAAACATTGTGGTGACAAAAATTTAAT 4854
Db 1262 ----- 1262
QY 4855 GAAATGACGACAAATATCTAATATTGAGTAAGATAAGTCAAAAGATTAGTAAG 4914
Db 1262 ----- 1262
QY 4915 GAGCAAAACACAGTGTCTATTAACTGCAAAACAAAGCTTAATAACTACATACTTG 4974
Db 1263 -----GlnValAlaSerGlnProGluValSerProThr----- 1273
QY 4975 AAGAAATGACGATGATGCTTTTCTCTCTGCGACTACGATATATGTTTCATGCA 5034
Db 1274 -----ThrLysAspVal----- 1278
QY 5035 TTGATGCAATTATACAGATCCAGAACTTAAGATGAATGGGTGCAAAAGATTG 5094
Db 1278 ----- 1278
QY 5095 ATGGAAGTGGCGCAACGGAAGGTACATTTGGGTCAATACTACAAAGAAAAAGAA 5154
Db 1278 ----- 1278
QY 5155 AAAGCAAAATTAACAAAGCTCGATGCGCAAAATTTCTTAAGAGTCCCGCTGTAGT 5214
Db 1279 -----LysProProCysAsp 1283
QY 5215 GCTAAGAAATATAGTTTATGATTTAAGATATATTAAGTATGATTAATTTGAA 5274
Db 1284 ILevalLys-----GluLeuPheSerAsnValAspThrLeuGln 1296
QY 5275 GATGAAAAACAAAGACCGAGAAATTTGAAGAAATATTAAACAAAAATGAAACATCA 5334
Db 1296 ----- 1296
QY 5335 GTTGGCAAGAGATGATAGTACTACAGAAATCCCGTACTGCGCGAAATTTTTC 5394
Db 1296 ----- 1296
QY 5395 TGAACGAAATTAAGAAATGTGTGGAACGATATGATCGGGTACAAACGTGTAGG 5454
Db 1296 ----- 1296
QY 5455 GATGATGAAATAGTGAATATGCAAGAAGTATGATGAATCTAATAAATGTTCT 5514
Db 1297 -----LysAlaCysSerThr 1301
QY 5515 GTACCTCAGATGATTAATTCCTATGGGAAAAATCGCATGAGGTACTGCGTATCAG 5574
Db 1302 -----LysTyrGlyProGlyLysAsn----- 1308
QY 5575 TTTCTTGATGTTTGGCGAATGGGTGAAGATTTTTCAAACATAAAGAAAGATTG 5634
Db 1308 ----- 1308
QY 5635 GAGAAATGGTAGGGCGGTGAATGATTAATCTTGTGATATAGTAAGATAAAGAAAG 5694
Db 1308 ----- 1308
QY 5695 AAATGACAGATGCTGTACACATATATAAATTTATTAGTAGAGGAACACAGATAT 5754
Db 1308 ----- 1308
QY 5755 GAAAAACAAATCAAAAAATATGTTGAGAAATTAAGCAAAAAATATATTCGAGATCTGTG 5814
Db 1309 -----TyrGly----- 1310

QY 5815 GCAAAAGATCGACAGACGCTCCGGAATATTAGACAAACATTTAAAAATTGTGAA 5874
Db 1310 ----- 1310
QY 5875 AATAAAGTGAGATGTGAATATAGGTATGAAAGATGTGCCACAGACGATTAAT 5934
Db 1311 -----TPArgCysIleProThrIysThrSerAsnAspValThr 1333
QY 5935 GATGTAATAGTCAAAATATGCCCCATCATTAGACGATGACCAAAAGATTGAAGA 5994
Db 1324 G1yGluAspG1yGlu ----- 1328
QY 5995 AAGTGAATTTGTCAAGTCCACAGAGTCCACAGCTGACGAAGGAAACCGCTACCA 6054
Db 1329 -----G1ySerArgValValArgSerThrProGlu 1339
QY 6055 CGGGTATCACTGATATCAAAAGGAGCATCGAAAAAAGAGAAACAGCGCGCT 6114
Db 1340 -----SerG1ySerAsnSerAspLysAsnG1yAlaThrCysIlePro 1354
QY 6115 ACAAAACAGCCGAAAGAAAGTGAATAACAACAGAAATGCGAGCACAACAGAAC 6174
Db 1355 ArgArgArgArgLeuValG1yLysLeu ----- 1364
QY 6175 CGACGACGACACACACACACACACATCAACACACACACACACACATTCGAC 6234
Db 1365 -----G1yGluThrAlaAsnLysHisAsnThrGluThrSerValSerGlu 1380
QY 6235 GTGGGCACATGGTAAAGCCATCTTCGAAATTAACCATAGACAGGGTGGAAATGAG 6294
Db 1381 G1uAlaThr -----G1uAlaArgG1y 1387
QY 6295 GGTGTAAATCCAAAACGATATGACATATCTAAATGGGTTGATATGATAGTACT 6354
Db 1388 -----SerGluAlaProAlaProGly -----G1y 1395
QY 6355 AAGAAATATGAAATGCGATATGATCCCTCCAGAGAAAAAATATATATATATAT 6414
Db 1396 LysGluSerSerGly ----- 1401
QY 6415 ATCAATATTTAAATATGAACATGAATAATAGCGTACATGATATAAAGAGCTTT 6474
Db 1402 -----G1yLysGluThrProSerAspLys -----LeuArgThrAlaPhe 1414
QY 6475 ATTAATATGTCAGCATAGAACTCAATTTTGTGTTAAATATATATATATATAT 6528
Db 1415 I1eGluSerAlaAlaValG1uThrPheLeuThrAspArgTyLysLysGluThrPleu 1434
6528 ----- 6528
Db 1435 Alag1uLysLysAlaGluLeuGlnAsnG1yG1yLeuAspLeuTySerSerG1yAspG1y 1454
QY 6529 -----AATCGTCGACGAGAAATGAAATGCAAAATGCAACATTCAGATGAATTT 6579
Db 1455 AspProAspAsnPro -----GlnAsnLysLeuLeuAsnG1yValIleProProAspPhe 1472
QY 6580 AAAAGAAATTAAT 6639
Db 1473 LeuArgLeuLeuPheTyThrLeuG1yAspTyArgAspIleLeuValHisG1yLys 1492
QY 6640 TCTAATCATATAAAAT 6699
Db 1493 ThrSerAspSerG1yAsnThrAsnG1ySerAsnAsnAsnIleValIleGluAlaSer 1512
QY 6700 AATAAGAAAAACAGATATAAAAGAAATGAGAAATATACGTAATAATA 6747
Db 1513 G1yAsnLysGluAspMetGluLysI1eGluGluLysI1eGluGluLysLeuProLysAsn 1532
QY 6748 -----TTTGGGAGAAA 6759
Db 1533 G1yG1yThrProLeuValProLysSerSerAlaGluThrProAspLysTrpTrpAsnGlu 1552
QY 6760 AATAAAAAAATTTATTTGGAGAAAGATCATATATATATATATATATATATAT 6804

Db 1553 HisAlaGluSerIleThrPlyG1yMetI1eCysAlaLeuThrTyThrGluLysAsnPro 1572
QY 6805 -----CATCTCACAGACGAAAGAAACGAAAGAAATTTAGACATTAATTTAC 6849
Db 1573 AspThrSerAlaArgG1yAspGluAsnLysI1eGluLysAspAspGluValTyGluLys 1592
QY 6849 ----- 6849
Db 1593 PhePheG1ySerThrAlaAspLysHisG1yThrAlaSerThrProThrG1yThrTyLys 1612
QY 6850 -----CAGTACAT---GACATGACCAACTG----- 6873
Db 1613 ThrGluThrAspTyG1yGluLysValLysLeuGluAspThrSerG1yAlaLysThrProSer 1632
QY 6874 -----ACGCCCTCCCTGAGAGAGTTGTATAAAAGCCCCCAATTTTGAGATG 6921
Db 1633 AlaSerSerAspThrProLeuLeuSerAspPheValLeuArgProProTyPheArgTy 1652
QY 6922 TTCACAGAAATGGCGACAGAAATTTTGTAAATAGAGAGAAAGAGTGTAAATTTGAG 6981
Db 1653 LeuGluGluThrPglGluAsnPheCysLysGluArgLysLysArgLeuLysGluI1eLys 1672
QY 6982 GCGGCGCTGT-----AACGAATATGAGTGAATGCT----- 7011
Db 1673 G1uGluCysMetAspG1ySerAspLysLysTySerG1yAspG1yGluGluCysAspArg 1692
QY 7012 -----AGTAATGACGGTAAAGACACAGAAAGATGCGAGAGCGGT 7050
Db 1693 ArgAspThrSerAsnGluValSerAlaAspLeuGluArgTySerCysG1yAsnSerCys 1712
QY 7051 GTAACATATCAAAATTTTATTAAGAAAGTGAAGAACTGAATATGAAGACAAAGAAAG 7110
Db 1713 ArgPheTyLysLysTrpI1eLysArgLysArgG1yGluTyAspLysGluAlaAsnAla 1732
QY 7111 TTCAAAAGCATAAAGATGCGAAAAATATATATATATATATATATATATATATATAT 7170
Db 1733 TySerLysGluLys-----ThrLysTyGlu----- 1741
QY 7171 GAGAAAGCAACATGCTGCTCATGAAATATTTAAACATGAATAATTAAGAAATTTGTGCAAT 7230
Db 1741 ----- 1741
QY 7231 AAGGATTTGCTGTATGACAAAACCTTCTTCACACTACCAAAACACAAATCA 7290
Db 1742 -----G1uG1ySer 1744
QY 7291 CAATCATCCGATGCTATATGATATGCAAGATCGCTGATATATGCTCGAAGATTTTAC 7350
Db 1745 LysG1yAlaG1yLeuAsnAsp-----HisAsn 1753
QY 7351 AAGGTGAGTGTCTGCTGAACCTTCAAAAAGGATCTATGATTCATACAAAAAATTTACT 7410
Db 1754 LysGluPheCysValLysLeuG1y----- 1761
QY 7411 GAACCTAAATATCCTATGAATTTGTAGAGAAAGACCATATTAAT----- 7455
Db 1762 -----ThrCysThrAspAlaI1aI1aPheLeuAsnArgLeuLysAsn 1775
QY 7456 -----TTACTTAAGACGAGAAAT---AATATGATTTTACTCTG---AAG 7497
Db 1776 G1yProCysLysLysAspAsnGluAsnG1yG1yAsnAspI1eAsnPheG1yAsnThrGlu 1795
QY 7498 GAAAAATTTTATCTATGATGCTACAAAGAAAGAAAGAAAGTAAATATATGTTGACTAAT 7557
Db 1796 G1uThrPheArgProAlaGluAsnG1yLys----- 1805
QY 7558 AATATCTTTCGATCTTAAGAAACCTTATGACCTGATATAATATATAGAGAGAGAAC 7617
Db 1806 -----ProCysSerSerPheLys-----I1eAsnCysArgAsn 1816
QY 7618 ---CCTTGTAAGAAATATAGAGAAAGAAATCGTTTAAAGTATGATATGAGAAATGTAC 7674

Db 1817 GlyAsnGlySerGlyAspGly-----AspThrLysGluLysCys--- 1830
QY 7675 AAAAATTCAGTCTATCAGAGAAAAAGATATGTGTCCTCCAGAGAACAT 7734
Db 1830 ----- 1830
QY 7735 ATGTCTTAAGCAATTTAGATGAATAAATTGAAGACTTAAGATAGTAATTATCTC 7794
Db 1830 ----- 1830
QY 7795 CTAAAAATGCTCTCGAAGTCGACGAATGAGATAGCATATATAAACTTCAAC 7854
Db 1831 ----- 1831
QY 7855 TCAGGAACGGTGGCGCAATGATTCATATGTGATCTATGAATAATAGTTGCGTAT 7914
Db 1842 ThrMetGlyThrCysThr----- 1847
QY 7915 CTGGGTGACATAGTTAGAGGAACAGATATGTTACGAATTTGGTGTACTTACCTCCGCTA 7974
Db 1848 ---GluAspValVal----- 1851
QY 7975 GAAATTAATTAATTAAGTTTGAATACATATATGGAATAAGAGAAATTAATAA 8034
Db 1852 ----- 1852
QY 8035 GGTAGAAATTAATTAACACGATGTACAAACGTTCTGCTGCTGCGATGCTAATGA 8094
Db 1862 PheGluGlyAspGlyLeuAspGluAlaCysGluAsnAlaGlyIlePheThrGlyIleArg 1881
QY 8095 AAACATATTTGGAAGCAATGACGTGCAAGCAACCAAGATGCAAACTTTTAGAAA 8154
Db 1882 LysAspGluThrLys-----CysArgLysValCysGlyLeuHisIleCysLysGlu 1898
QY 8155 GGAAGATGATGATTT-----GAACGATTAACATTAACAGATTAAGTGTGCA 8205
Db 1899 GluLysGlyAsnGlyAlaIleAsnAspGluGlnIleIleLeuValArg----- 1914
QY 8206 CATAAAGCAGATCCACCTGTTGATGATTATATCTCAACGTTGCGATGACTGAA 8265
Db 1915 ----- 1915
QY 8266 TGCTGCAATATTAATTTAAAGCAGTGAAGAGAAATTTGAAAAATTAATCATGT 8325
Db 1920 TrpValGluThrPhe-----LeuGluAspTyrLysIleLysLysLysLeu 1935
QY 8326 GATCACTGTAAACATCTGACAGATGCAAGATGATTATGATGAATTAAGTGAACAG 8385
Db 1936 LysProCys-----IleGluAsnGlyAsnGlySerThr 1946
QY 8386 TGTAAACAGATGTCAGAAATATAAAATTTGTTCTTAATGGAATCTCTATTGAT 8445
Db 1947 CysIleAsnGlyCysAsnLysLysCysAsnArgValGlyIleLysLeuLysLys 1966
QY 8446 ATACATCAAAATTAACAAAGATTTAT----- 8475
Db 1967 AspGluThrPheLysIleLysAsnHisTyrLeuGluLysAsnGlyLysAspLysAsn 1986
QY 8476 -----GAACACCAATATATATCAAAATCTCTACTATGAT 8511
Db 1987 ValThrSerLeuValThrAsnValLeuGluThrLeuValThrGlnIleAlaIleAsn 2006
QY 8512 -----CATGTTCAAAATTTGTACAAAAGTTGAANAAT-----TTTAAAGATGAATGCTCT 8562
Db 2007 AspLysArgGluGlnThrSerLeuAspLysLeuLysThrSerLeuGlyCysAsnGlySerPro 2026
QY 8563 GTTGAGACCTTTTCGATATCTTCATGAACACAGTAAGTGTGGAATTAATTTAAT 8622
Db 2027 GluAsnSer-----Arg 2030
QY 8623 GAAATGATGCTCTTCCAAATATGCAACATATGCTTTCGAGAAGAACACCAAAAGTTAT 8662
Db 2031 LysAsnAspGlyAsnGluAsn----- 2037

QY 8683 AAAGACCTTCAGATGTTGTAACCTTCTTAAGATTCATGATGATATATGCTACCGAT 8742
Db 2038 ---AspAlaIleAspCysMetLeu----- 2044
QY 8743 CAAAACAAAGATGATGATTAAGAAATTAACAACTTTACCTTCTGCTCGAAGATGATAT 8802
Db 2045 -----AsnLysLeuGluThrLysIleHisGluCysLysThrGlnHis 2058
QY 8803 GATTAATATCTGATATATGTAATGCAACGATACCTGTTCTTAATAGTTCAATGAT----- 8856
Db 2059 GluAsnSerValGluAsn-----SerAspLysProHis 2069
QY 8857 ---AACAAAGCTGTATGATTCCTCCAGAGAAGACATTTATGTACAGACCTATCACT 8913
Db 2070 ProAsnGlyGlyLysAsnProPro----- 2078
QY 8914 GCATATATTAATTAAGAAAGTGATTAAGAAATTTTAAAAAAACCTTCTACTTCTGCT 8973
Db 2078 ----- 2078
QY 8974 TTCAGTCAAGCAATTTGTTAGGTCAAAAATATATAATGGAAGAGAGTGTGCTTGAG 9033
Db 2079 -----AspGluGluAspLeuLeuGln 2086
QY 9034 GCAATGAATATAGTTATGTCACATTAATTCCTCGATATTAATTAAGAACTGATATGATGAC 9093
Db 2087 Glu----- 2087
QY 9094 ACTTCATATCTGAATAAATTAATAAATATTGAAACATCAATGAACCAACCGAAAT 9153
Db 2087 ----- 2087
QY 9154 CGTAAACATGTTGGGAATTAATATAGCTGATGACGAGTATGATGTGATAT 9213
Db 2088 -----GluAsnProValGluGln----- 2093
QY 9214 AAAATGCTACTTCAAAGTAACATTAAGATGAGAGTGTGCAATTAACCAAGATGAA 9273
Db 2094 -----ProGlyPheCysProThrProGlnGlnGlu 2103
QY 9274 GAACTAATCACTTCTCTGTTGTTAATGAATGGCAAGCAGCATGAAGAAAG 9333
Db 2104 ProGluProAsp----- 2107
QY 9334 AAACATGTAAGTATGATTCATTAATAAACAATGCTCTGTTCAACGAAGTAAATTTGAA 9393
Db 2108 -----AspLysCysGlyLysLeuGluGluLysLysAspGlu 2119
QY 9394 GCGTCAGATTTAATGAACAACCTGGATGTCAGATGATATTAAGAAATATATTACTG 9453
Db 2120 LysLysGlu-----GlnPro----- 2124
QY 9454 AATATATGTAATAAATTAACAAATGAAGAAATCTAAATTAATAAATTAAGCAATTAAGAT 9513
Db 2124 ----- 2124
QY 9514 CAATCTCAGTAATATAGACATTAACCATCTGAAGAAATGTTCAAGCATATATAA 9573
Db 2125 -----GluGlnProAlaGluGlnAspGlyGlyAlaIleValPro 2137
QY 9574 TCAAAAGATTTCTCAATGCGCTTTGAGTTAAATGATATTAATGAAATAGTTACAGAGACA 9633
Db 2138 SerGlyPro----- 2140
QY 9634 AAAAATTAATGAATAATATGATTCACAAGAAGTACTAAAAAATTAATATCTGCTTAAAT 9693
Db 2141 -----ProGly----- 2142
QY 9694 TTTGTTGAAGTGAAGAACACAAATATCATGTACTAGATGAATAATATAAAGAAAGAG 9753
Db 2143 ---SerGluProGluLysAspLysGlyProValLysProAlaGluIle----- 2157

```

QY 9754 CAACAGTGTGCTTAAGACATCTATTCTTACACCCATGATGTTCTTCTATCAA 9813
    |||||
Db 2158 -----ProlysPro-----Gln 2161
QY 9814 GCACCTTATTCACACATCGAGTAGCACATATGATCTTAATAATGATATATGAA 9873
    |||
Db 2162 GlutroProAspLeuSerHisProAla-----ValIlePro 2173
QY 9874 AGTATVCTCTGTGTGTTGATCGCGCTAGGTTTG-----ATAGCGCTT 9921
    |||
Db 2174 SerLeuValHisSerThrLeuAlaIlePbSerValGlyIleGlyPheAlaIlePheThr 2193
QY 9922 CATTTCAATGAAGAAAAATTCAAATGCTGTG---GACTGTGTCGTATATGATATC 9978
    |||
Db 2194 PheTyrLeuLysLysLysThrLysSerSerValGlyAsnLeuPheGlnIleLeuGlnIle 2213
QY 9979 CCGCAGAGAGATGATGAGATCGCTAGTTCGAATCCAAATAGTTCATACCATATAGA 10038
    |||
Db 2214 ProlSerAspTyrHisPheProThrLysLeuSerProAsnArgTyrIleProTyrThr 2233
    |||
QY 10039 AGTGTCCATATTAAGCAAAACATATATATATATATGAGAGAGATCT-----AGT 10089
    |||
Db 2234 SerGlyLysTyrArgGlyLysArgTyrIleTyrLeuGlnLysPheSerGlyThrAspSer 2253
QY 10090 GGAGATGACATATATATATATGCTGACTTATCTCTGATATATCTATCCGAAAGT 10149
    |||
Db 2254 GlyTyrThrAspHisTyr-----SerAspIleThrSerSerGlnSer 2267
QY 10150 GATATGATGAAGATTTGATATATATATATATATATATATATATATATATATATAT 10209
    |||
Db 2268 GlnTyrGlnLysMetAspIleAsnAspIleTyrValProLysSerProLysTyrLysThr 2287
QY 10210 TTGATAGACTAGTACTAGAACCAATCAAAAGGATATATACCAAGTGTAT----- 10257
    |||
Db 2288 LeuIleGlnValIleValLeuGlnProSerGlyAsnAsnThrThrAlaSerGlyAsnAsnThr 2307
QY 10258 -----GATACACCAACT----- 10269
Db 2308 ThrAlaSerGlyAsnAsnThrThrAlaSerGlyLysAsnThrProSerAspThrGlnAsn 2327
QY 10270 -----AATGATACACACAGCAGCAATAGATTTATGATGATGATGATGATGATG 10320
    |||
Db 2328 AspIleGlnAsnAspGlyIleProSerSerLysIleThrAspAsnGlnIlePheGlnIle 2347
QY 10321 AACCATGATTTTGTATCTCAATATTTA-----CCAAATACAGAACCAATATATAT 10371
    |||
Db 2348 LysAspGlnPheIleSerGlnTyrLeuGlnSerGlnProAsnThrGlnPro-----Asn 2365
    |||
QY 10372 TACAAAGTCAGATATTCACATGATACAGAACCTTAATCTTATATCTGATATACCT 10431
    |||
Db 2366 MetLeuGlnTyrHisValAspAsnAsnThrHisProThrThrSerHis---HisAsnVal 2384
QY 10432 GAAGAAAAACCTTTATATATATATATATATATATATATATATATATATATATAT 10491
    |||
Db 2385 GlnGlnLysProPheIleMetSerIleHisAspArgHisLeuPheSerGlyGlnGlnTyr 2404
QY 10492 AGTTAT-----AATATTAATATAGACTATAAT-- 10518
    |||
Db 2405 AsnTyrAspMetPheAsnSerGlyAsnAsnProIleAsnIleSerAspSerThrAsnSer 2424
QY 10519 -----ACTAATATATGATATTCATGATGATGATGATGATGATGATGATGATG 10563
    |||
Db 2425 MetAspSerLeuThrSerAsnAsnHisSerProTyrAsnAspLysAsnAspLeuTyrSer 2444
QY 10564 GGATATAGATTTAATATATATATATATATATATATATATATATATATATATAT 10623
    |||
Db 2445 GlyIleAspLeuIleAsnAspAlaLeu--SerGlyAsnHisIleAspIleTyrAspLeu 2464
QY 10624 TATTTG 10628
    |||
Db 2464 etLeu 2465

```

```

T28157
erythrocyte membrane protein 1 - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C/Accession: T28157
R/Tang, S.Q.; Wollish, W.S.; Gut, J.; Wu, J.; Ahn, J.; Petersen, C.; Fujioaka, H.; Aik
submitted to the EMBL Data Library, July 1995
A/Description: The molecular cloning and DNA sequence analysis of Plasmodium falcipar
A/Reference number: 220479
A/Accession: T28157
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-2212 <YAN>
A/Cross-references: EMBL:U01083; NID:g1517813; PID:g1517814; PIDN:AAB06961.1
C/Genetics:
A/Gene: EMP1

Alignment Scores:
Pred. No.: 2,22e-124 Length: 2212
Score: 2678.50 Matches: 719
Percent Similarity: 44.45% Conservative: 311
Best Local Similarity: 31.03% Mismatches: 809
Query Match: 13.77% Indels: 478
Gaps: 84

US-10-087-013-1 (1-10628) x T28157 (1-2212)
QY 91 AGTCACAAAGTCGAGAAATGTTTGGACGTTATGCAAAATATAGACATCATCA 150
    |||
Db 9 ThrAsnLysSerAlaLysGlnValLeuAspGlnIleGlyIleThrIleGln----- 25
QY 151 AATATATGCAAAAGACATGCGATTCG-----TTGAAAGGGATTTG 192
    |||
Db 26 -----LysLysAlaHisSerAspAlaAspThrPheArgSerGlnLeuLysGlyAsn 43
QY 193 ACAGAACAGATTTCTGCGTGT-----CCTTACGCCAGTAAATAG 237
    |||
Db 44 GlyIleAlaLysPheTyrAsnGlyGlyIleMetGlnProAsnSerLysLeu----- 61
QY 238 CATATATATATATATATATATATATATATATATATATATATATATATATATATAT 297
    |||
Db 62 -----CysGlnLeuAspHisThrIleAspThrAsnValThr 73
QY 298 TATGATATGTAATTTGACATCTTCCATGATAGACAAACCGATTTGATGA 357
    |||
Db 74 AspGlyHisSerAsn-----ProCysGlnGlyArgGlnThrValArgPheProAsp 90
QY 358 GATGAGAACTCGAATGT-----GGAATTAATATAGTATTTAATAAGAAAAATGATGCT 414
    |||
Db 91 AspAsnArgSerGlnCysThrLysAsnArgIleLysAsp---SerValAspAsnSerVal 109
QY 415 ATACCTGTGCGCCACCTAGAACGACATATGCTGATATAAACTTGAAGCTCTAAAT 474
    |||
Db 110 GlyLysAlaLysPheTyrArgArgLeuHisLeuLysSerHisAsnLeuGlnSerIleGln 129
QY 475 GATATAAT-----ACCAAAATATTCATGATTTATGGAATATGTAAGTATCA 525
    |||
Db 130 ThrAsnAsnTyrAspSerSerLysAlaLysHisAsnLeuAlaGlnValCysTyrAla 149
QY 526 GCAAATATGCAAGGTGATGCAATGTTAATATATATCA-----CAATAAGAACT 576
    |||
Db 150 AlaLysPheGlnGlyGlnSerIleValLysAsnTyrGlnLeuGlnHisThrThr 169
QY 577 TCAGACCTGTACTGCTTTCAGCAAGATTTTCAGATATAGCTGATATGTAAGAGA 636
    |||
Db 170 GlnGlyIleCysThrAlaLeuAlaIleArgSerPheAlaAspIleGlyAspIleIleArgGly 189
QY 637 ATAGATATGTTTAAACCAATGTCATGAC-----AAATAGAAAACGGGT 681
    |||
Db 190 LysAspLeuTyrLeuGlnLysAsnProGlnGlnSerAlaArgArgLysGlnLeuGlnAspAsn 209
QY 682 CTCGAGAGGTTTCAAGAAATATACAT----- 708

```

RESULT 8


```
Db 210 LeuArgLysIlePheAspPheIleTyrLysGluLeuThrSerSerArgAsnGlyLysThr 229
OY 709 GATGGAATGGAAGATGAAGATAAATAATGATTACATCCGTGATGGATCGAAATATATAT 768
Db 220 AsnGlyAlaGluGluArgTyrLys-----AspGlySerGlyAsnTyrTyr 244
OY 769 AAATTAACAGAACGCTGGTGGATGTGATGCAATGAATAAATAGTAGGAAGCTATACATGT 828
Db 245 LysLeuArgGluAspTyrTrpAsnAlaAsnArgLeuAspIleTyrPheLysAlaMetIleCys 264
OY 829 GATGATCATATTAATCTCGATATTTTATGCAATCAGAAAGATATACACCTATTTTCA 888
Db 265 LysAlaProGlyAsnAlaGlnTyrPheArgAsnThrCysSerAsnGlyGluLysProThr 284
OY 889 AATCTTAATCGGGCCATTAACAGAAAGGCTCTACCAATTTAGATTATGTCCTCA 948
Db 285 GlyLeuLysCysGlnCysIleAspGlyThrValProThrAsnLeuAspTyrValProGln 304
OY 949 TATTACGTTGGTTCGACGAATGGGAGAGAGATTGTCGGAAAGAAATATTAATTTG 1008
Db 305 TyrLeuArgTyrPheGluGlnTyrPheAlaGluGluPheCysArgLysAsnLeuLysLeu 324
OY 1009 AAAAAAGTCAAGAGACTCTGTCTGT-----AATGACAAGACGCTATATTTAGT 1059
Db 325 GluAsnAlaIleLysAsnCysHisGlyMetAspAspArgLysGluLysTyrCysSer 344
OY 1060 CATTAATGAGATGATGTTGATGACACATATTGCAAAAAAGATTTTGCATTGATAT 1119,
Db 345 ArgAsnGlyTyrAspCysThrLysThrIleArgSerIleAspLysTyrSerMetAsnArg 364
OY 1120 AAGTGTACTGACTGTTGCAATAATGCAAAAGTTTGTGAAGTTGGTAGGAGATCAACA 1179
Db 365 GluCysThrLysCysLeuTyrValCysAspProTyrValLysTrpIleAspAsnLysLys 384
OY 1180 GAAGCATTTAAAAACAAAAAAGAAATATGAAGAAATACATCATATTATTGATG 1239
Db 385 LysGluPheGluLysGlnLysLysLysCysGluAsnGlnIleTyrArgAsnAsnGlnSer 404
OY 1240 GATACAAATTTGTCATATATATTATGATGATTTATTAACAATTTATGAAAACCT 1299
Db 405 SerGluAsnSerProLysAsnTyrAsnAsnMetTyrGlnThrAspPheTyrGluAsnLeu 424
OY 1300 AAGGAACCGCAATATGCAACATAAGCACTTTTAAATTTACTAATGAAGGAAGTAT 1359
Db 425 LysLysAspTyrGlnIleAspLeuLysAsp-----PheLeuLysLeuAsnSerGluThrPro 443
OY 1360 TGTAAAGAGAGATTCACAGAGAAAGATTTACTTTACTTAACAGTCTGATGACA 1419
Db 444 CysThrAsnIleIleAspLeuLysArgLysIleAspPheThrLys-----AspProGln 461
OY 1420 GGGATATTTTATCGTTGACAATATTCGCAAGTGTGCCGACTGGGGGTCAAA----- 1473
Db 462 GluThrPheSerHisThrGluTyrCysAspProCysProTyrCysGlyLeuLysLysGln 481
OY 1474 TGTGATGGATTAATAATACACACAAATACAGATATGATCGTAACGCTGAATATGA 1533
Db 482 AlaAspGlyLys-----ThrTyrAsnArgLeuTyrGluAsnAspProGlnCysProIleLysPro 500
OY 1534 GACTATTAACCTCATGGGTGTGAAGCTTACTATATCATCTGCTTATATAGTGTAT 1593
Db 501 LysTyrGluProProLysGlyGluLysProThrGluIleAspValLeuTyrThrGlyLys 520
OY 1594 GAACAAGGTGATATTACACAAAAATTTTGTACAGAGCTCAACTAATTTATACAA 1653
Db 521 GluAsnLysAspIleIleValLysLeuArgLysPheCysLysThrAspGlyAsnThrCys 540
OY 1654 GATAAATAATTCAAAAATGGGAATGCATATATAGATGTAAGTAATATTAATATGA 1713
Db 541 PheLys-----AsnGluGluTyrAsnCysTyrTyrGluValGlyAsn-----AspLysCysVal 558
OY 1714 CTGAACAAAAATACTGAATCATATATGATATCTTAAGATATATATCATATATTT 1773
Db 559 LeuGluAsnGlyGluGluLeuGlyGlyLysLys-----LysValLysAspTyrAspAsnPhe 577

OY 1774 TTGGAATATGGTTACATTTTATTAAGGATACATTAAGTGAATGACAACCTTAA 1833
Db 578 LeuMetPheThrValAlaHisMetLeuLysAspSerIleGluThrArgSerLysLeuSer 597
OY 1834 ACTGTATTAATTAATCAACCCAGCATTTGATGATGATGATGATGATGATGATGAT 1893
Db 598 AsnCysLeuLysSerAspLysLysThrCysIleLysLysCysAsnAsnAsnCysCys 617
OY 1894 TTTGACAGATGGGTAAACAAAAAGAAAGAAAGATGAAATAGTAAAGAAAGCTTCA 1953
Db 618 TyrGluLysThrIleGlyLysLysValGluThrPheGlnIleLysLysHisPheAsp 637
OY 1954 AAAAAAAGATATACAG-----CAATGCTATTAATGATATTAATATATCTTTGAAGT 2010
Db 638 LysGlnThrAspPheGlnIleTyrGluArgTyrPheValLeuGlnThrValLeuGlnGly 657
OY 2011 -----TATTTTATTAAGTTATGATTAACCTTGACAAGATGACAAAGATGCAAAAGACT 2067
Db 658 AspGlnPhePheThrAspIleThrLysAlaTyrGlyAspAlaArgGluIleValHisIle 677
OY 2068 ATGGAATATATTAATAAAGAAAAAATAGCTTTTCCATTTGAAATATAGGAGTAT 2127
Db 678 GlnGluMetLeuGlnLysLysLysGluGlnVal-----LeuHisGluAspAlaSerAsn 695
OY 2128 TTAGGAATGCAATGCACTGTTGATGATCACTTAAAGAAAGCTCCACGATATGT--- 2184
Db 696 MetLysThrIleIleAspLeuLeuLeuAspHisGluLeuLysGluAlaLysGlnCysIle 715
OY 2185 -----AAGACAAATATTAACAACGAGCATGCAACCTCCCAAT----- 2226
Db 716 ValAsnHisLysAspAsnAsnCysProAspAspLeuSerAspSerGluAspGlnGlu 735
OY 2227 -----GCAACAACAAACCCGTGTGTAACCTCTGAGGACGCAACCACTAA 2277
Db 736 AspIleProGlnArgGlnAsnProCysAlaLysProSerGlySerAlaHisArgAlaLeu 755
OY 2278 AATATTAAGAAATAGCAACATACTTTAAAGGATGATGAGGAGCAACGAATCGT 2337
Db 756 ---ValAsnLysValAlaSerAsnMetHisLysLysLysArgGlnLeuValAsnAsp 774
OY 2338 GGTCTT---CATTAATTAAGAAAGCAACAGCAACAGGTATATTAACGTTGGGTAGG 2394
Db 775 GlyValSerSerLysLysLysGlyAspAlaAlaLysGlyGluTyrArgLysSerGlyThr 794
OY 2395 AGAAAGACTTCAAGCAATTTATGATGATTAATGATTAACCTTATATCGTAATCTT 2454
Db 795 ThrIleLysLeuLysAsp-----IleCysSerIleThrAsnAspHisSerAsnAlaLysArg 813
OY 2455 GGTTTTCAATGCAACCATGATGATGCAAGGCAACAGGTGATATACAAAGATTT 2514
Db 814 GlyHisThrAspGlnProCysLysArgLys-----AspSerLysSerGluMetPhe 830
OY 2515 GTCTAGGAATGAAATGAGAGTGATCCGGAACACATCGGTAAAGATCAACGAATGTT 2574
Db 831 ArgThrGluAspGlyTyrLys-----ProAlaGlyPheIleSerLysThrTyrLysAspIle 849
OY 2575 ATTATGCTCTTACAGACAGCATATATGATCCAAATTGGAACATTTTACAAAGGAT 2634
Db 850 TyrMetProProArgGlnHisPheCysThrSerAsnLeuGluTyrLeuGlnThr 869
OY 2635 GATCACCCTTAATGATATATTTGTGAT-----GATTAGCTTAATAATTCCTTT 2685
Db 870 AsnLysLeuLeuAsnGlyAsnAspIleAsnGlyAsnProAsnIleIleAsnAspSerPhe 889
OY 2686 TTGGGGATGTTCTTCAAGCAAAATATGAGCAAAAGATTAATGAAATGATGATTA 2745
Db 890 LeuGlyAspValLeuPheAlaHisTyrGlnAlaAspPheIleLysLysMetTyrLys 909
OY 2746 GAAAAGATTAACCTAAAGGGCCCAAGAAAGTAAGTACGCAACCAACGACGATATC 2805
Db 910 LysGlnAsn-----AspTyrLysAspAsnAlaThrIle 920
```

OY	2806	TCGCGACGTATACGCTACGATTTTGGAGATATAGGTGATTAATTCGAGGAGAAATGTC	2865
Db	921	Cysatgataacetylserlystserphealaaspleinglyaspilleuleysgilythrinspmet	940
OY	2866	TGGGAAAGAAACGGGTGCATGTGTAAAGCGTCAAGGACATTTGGGAAACGTGTTTGGTAT	2925
Db	941	Trpaspnaaspserserglyuuserlystlnrthasrplyleuadrgulilepheasrphr	960
OY	2926	ATACATTAAGTCACTCAAAAGCCAAAGAAATGATTAATATATATGATGATGCCCAAAATAT	2985
Db	961	Ileuyslystlnrshisproglyle--lysগুলিত্যেতুল্যগুলিাasrphrpro--tyr	978
OY	2986	TTAAATTTGGGAGAAATTCGTGGGAAAGCTAATATAGAGCCAAAGTATGGGAAAGCCATGAA	3045
Db	979	Thrlusleuadrguliasrpttrprrglualaasnaarplyslyliletrpbulametein	998
OY	3046	TGTGATGTAAATATATTGAAGGATTAATATCGGACACCAATCAACAAAGTAGTATATGC	3105
Db	999	Cys-----prothirproasnglyserphero	1007
Db	3106	GGATATATGATCATATACACCAATTCGATTAATATATCCCAAAATTAAGATGATGACC	3165
Db	1008	Cyslysserlythlnsrthproleuadrasrplylileproglinargleuargtrpmetthr	1027
OY	3166	GAATGGCCAAATGTACTCTCCAAAGTGCAGAAAAGAGATATGATTAAGTTGAAGAGAG	3225
Db	1028	Gluutrpalaeutrrprrhecsylsuglnglnlylsylsnglnttylglyleuvalaserlia	1047
OY	3226	TGTAAAGAGTGTAAAGATTAATATGATTCGCAAGGCTGTACAAAGAGAGTGTACAGT	3285
Db	1048	Cysasnglycyslysasrglu-----glylsvalcyslthasnllyser---sergin	1064
OY	3286	TGTACGAAGTGCACAGACCTGTAAATGATTAATATGATTAATAGATTAATGAAAGAA	3345
Db	1065	Cyslthsercysmetglnalacysgulsanthylysasnrheileasnthrtpryglu	1084
OY	3346	CAATGGATATATATATGATGATTAATACAAAGATTCATGAAACAAAGCAAAATGCTGTT	3405
Db	1085	Glntrpasprrpysmetguluilelystlytlysleuiletytyleuglnlaiglnlthrtthria	1104
OY	3406	AGTAAATAGTGGTATTAAGACCTTCACAGTACGCCAAATATCATATAGACAGAAATGTAT	3465
Db	1105	Alasnglycylproasrphltyrserglyleuvalasrbulasnglulysrprovalayal	1124
OY	3466	GAATTTTTCGCGAATTAATACAAACAAATGCTGCGAAA-----	3504
Db	1125	Asnrheleupheglnleuetytlysgulsannglylylelglyasnrproalgsrphr	1144
OY	3505	-----AGTAAATAAACTGGTACTAGTATGAAGAACTGCTGTCAATTGGTACTACACACG	3558
Db	1145	ProatrgulalysasrserltsatrgulthrtharprolalaSerValalalysasnsrparal	1164
OY	3559	TATGAAATGTGGAGATATTCATTCATC-----GATACGSGAAATTTGATGATGTGCAGTCA	3615
Db	1165	Tytsertthralaaglytvalalnshnglnmetelyprohismetelsuylsyethr	1184
OY	3616	CAAAATGAGTTTTCGTATGATGAAAAAAGTCAATGTCAGATGAGATGAAAAAATGCTCTTGA	3675
Db	1185	Glnthrgluphecs---gluyslthasrbulnglntlyrasnbulasntlythrphelys	1203
OY	3676	GATTAACCAACAGCACATGATGCTGGCTGTGGTTGTAAAGTGGATCGAAACCGACAAAG	3735
Db	1204	Asnrproproglntlytlysasrpalacysllecysasnthargproproglyslu	1223
OY	3736	GTAACGATTAATAACGAAAAA-----AAACG	3762
Db	1224	Aspserrglysalrgsergluaspserrasrpglulglyvallysglulthrlysva	1243
OY	3763	GAAGAAAAGGATGAGAA-----	3780
Db	1244	GluGluysalatrthrgluasrpalavalasrpthrclpyrproproalalprolysglula	1263
OY	3781	-----TGTAAACAGTGCATGATATATCTTAAGAAAAACGATGAGAA	3822

Dd	1264	ThrThrLeuAspValCysProIleValAlaGlyValLeuThr-----Lys	1279
Qy	3823	AAACAGATAGAAAGATTGTCATCCAAAAAAGAAATGTAATGATATCC---GATTGGCAA	3879
Dd	1280	GLuAsnLeuGluAsnAlaCysProThrLysTyrGlyProLysAlaProThrSerThrLys	1299
Qy	3880	TGC-----	3891
Dd	1300	CysIleProThrGluLysThrAsnAlaAlaThrGlySerGluGlySerSerCylsasnGly	1319
Qy	3892	AATTATAGTGAAGACCTCGT-----	3912
Dd	1320	AlaLeuGlnAlaGalaLysAlaGalaThrValGluSerGlySerProValThrSerAsnSer	1339
Qy	3913	-----GTGTGTATGCCCCCTAGACAGCAAAAGATWATGCGTA-----CATTTCTTG	3957
Dd	1340	GlySerIleCysIleProProAlaArgAlaArgLeuTyrIleGlnLysLeuHisAspTrp	1359
Qy	3958	GCAAATGATTAATGAATTAATAAAAAATTACATCAAGCTT-----	3966
Dd	1360	AlaSerGlyAsnThrValValSerGlyGlnAlaGlnThrProGlnGlyCylThrSerSer	1379
Qy	3997	-----AATTAAAGAAAGCTTTCATCAATCTGCACA	4029
Dd	1380	ProSerGlyLysGluThrProSerAspLysLeuArgThrAlaPheIleGlnSerAlaIa	1399
Qy	4030	GCAGAAACATTCCTCCTCATGTATTAATTAATAAGTAACAGATGCAAGAAATCACTC	4089
Dd	1400	IleGluThrPhePheLeuThrPheAspGlyTyrLysLysGlu-----LysGluIle	1415
Qy	4090	GATAAAAGATTAAAAAGAA-----	4107
Dd	1416	GluLysLysGluLysLysValAlaAsnGlyGlyLeuValProSerLeuAsnGlyLysPro	1435
Qy	4108	-----GGCAAA	4113
Dd	1436	ProGlnGlnProGlyValThrGlyAspSerProGlnSerLysLeuGlnGlnThrGlyVal	1455
Qy	4114	ATTCCTCCCGCATTTTGTGAGATCCATGTCATCCATTTGGAGATTAATGACATTTTGTTA	4173
Dd	1456	IleProProProPheLeuArgGlnMetPheTyrThrLeuGlyAspTyrAlaAspIlePhe	1475
Qy	4174	TTTGGACACAT-----ATATCAAAAGCTCATGCTGAG-----	4206
Dd	1476	PheGlyLysAsnAspIleValIleAspThrLysAsnGlyAspLysAspIleAlaGluArg	1495
Qy	4207	GGAGTAAACTAAAAAGCAATAGATTCCTTTTAAAAATGATGACCAAAATCTCT	4266
Dd	1496	GluLysLysIleLysAspAlaIleGluArgValLeuLysAsnAlaAspSerGlnProPro	1515
Qy	4267	AATGCAAAACAGCCCAAGATGCTGGACAGCAATGACATGACGATATGGCAACCTATG	4326
Dd	1516	SerAspGluLysAlaGlnIleThrTyrProGluGlnAsnGlyGluHisIleThrAsnGlyLeu	1535
Qy	4337	CTATGTGCATAGTA-----AAAATT	4347
Dd	1536	IleCysAlaLeuThrTyrLysGluLysAspGluLysGlyThrProLeuLysGlnAsnGlu	1555
Qy	4348	GGGGCAAAA-----AAAGATGATTTTACGCAAAATACGCT	4383
Dd	1556	GlyLeuLysSerAlaLeuTrpAspGluLysAsnLysLysProLysAspLysIleLysTyrGln	1575
Qy	4384	TACAAACAAACCAATTATAGCAAAAGC-----	4413
Dd	1576	TyrAspLysValLysLeuAspGluAsnSerGlyThrSerProLysThrAsnAspHisVal	1595
Qy	4414	-----ACCACTTTGGAGAAATTTGCCAAACGACCCCATTTTATACATGCTATACGAA	4467
Dd	1596	ProProThrProLeuThrAsnPheIleSerArgProProTyrPheArgTyrLeuGlnGlu	1615
Qy	4468	TGTATCGACGACTTTCCTATACACGACAAAAATATTTAGAGATGTGCGAGGAAATATGT	4527

Db 1616 TrpGlyGluThrPheCysArgGluArgLysLysArgLeuGluLysLysValGluLys 1635
 QY 4528 AAGTCAAAATGAC-----CAATTGAAGTGTAT----- 4554
 Db 1636 MetaspGluAspGlyLysGlnLysCysSerGlyAspGlyLysAspGlyGluLys 1655
 QY 4555 -----ACAGATGTAATAGAAATGC 4575
 Db 1656 ArgLysGlnAspTyrSerThrValArgAspPheTyrCysProGluCysGlyLysTyrCys 1675
 QY 4576 GAGACTACGCTTAATATATGCAAAAAA-----GAGTGGATTCCACAAGATTAATAT 4632
 Db 1676 ArgPheTyrLysArgTyrPheGlnLysLysLysAspLysLysLysGlnLysGlnLys 1695
 QY 4633 TACAGAGATGACCGCAACAAAAAAGATTGATGACACACACATTGGTGAATGTTACA 4692
 Db 1696 TyrAsnGlnLysThrAspAlaArgArgAsnAsnAsnAspAsnAlaPheSerThrThr 1715
 QY 4693 GACTATACGTGACGACGACGACGATTTACTTGAACAGAAATTTTCTGCTACTGCT 4749
 Db 1716 LeuAspThrCysThrThrAlaGlyAspPheLeuGlnThrLeuLysAsnGlyProCysLys 1735
 QY 4750 -----GGTATTAAGCTCGAAGTCCCTGCTGTGACAAAGAAATATACAAATG 4797
 Db 1736 AsnAspAsnValAspAspSerGlyGluAsnLysLysLysPheAspGluAsnGlyAspThr 1755
 QY 4798 TTAGAAAAACAGGCTTACTATGATGCGACAAACATTGTGGCGCAAAATTTATTGAA 4857
 Db 1756 PheLysTyrThrGlnTyrCysGlyThr-----CysSerLeuAsnGlyPheLysCys 1772
 QY 4858 AATGACACAAA-----TATACATTTGCGAGTAAAGATTAAGTGAAGGATTA 4908
 Db 1773 AsnGlyAspAspCysArgValArgThrAsnValThr-----CysAsnGlySer 1788
 QY 4909 GTAAGAGAGCAAAACAGGCTTATTAGTGCACAAAGAGTCTTAATATCAAT 4968
 Db 1789 AsnArgThrThrThrIleThrIleAspAspPheLysAsnGlyLys-----AsnSerAla 1806
 QY 4969 AACTTAAGAATGATGACGAGATGCTTTTCTCTCTGCGACATGATATGTTT 5028
 Db 1807 GluIleAsnMetLeuValSerAsp----- 1814
 QY 5029 CATGCAATGCGCAATATATACAGTCCAGATTTAAAGATGAAATGGTTCGCAAAA 5088
 Db 1815 -----AspIleAsnSerGlyAsnGlyPheAsn--- 1823
 QY 5089 AGATTGATGAGTGGCGCAACGAGGATCAATTTGGGCAATACATACAAAGAAAA 5148
 Db 1824 -----AspLeuGlnLysCysLysAsnAlaAsnIlePheLysGlyLysLysGlnAsn 1840
 QY 5149 AAGAAAAAGAAAAATAAAAAGTCGATGCGCACAAAATTTCTTATGAGGTCCCGCT 5208
 Db 1841 LysTyrLys-----CysValTyrPheCysLysSerAspAla 1852
 QY 5209 TGTAGTGCATGAATATATGTTTATGATTTAAGAGATATATCTA----- 5256
 Db 1853 CysGlyLeuLysLysAsnAsnAspIleAspGlnAsnGlnIleIleLeuIleArgAlaLeu 1872
 QY 5257 -----GGTATGATTAATTTGGAAGATGAAAAACAAAGACCGAGAAATTTGAG 5307
 Db 1873 PheLysArgTyrPheGlnTyrPheLeuAspAspTyrAsnLysIleArgLysLysLeuAsn 1892
 QY 5308 AAAAATTTTAAACAAAAATGGAACATCACTTGGCAAAAGAGATGATAGTACAGAAAT 5367
 Db 1893 ProCysIleAsnAsn-----GlyGlnLysAlaIleCysThrAsnGlyLys 1907
 QY 5368 CCCGCTAGTACTGCGCAAAATTTCTGGAACGAAAAATAGATGCTGTGCAACCA 5427
 Db 1908 ValGluGln-----TyrLeuAsnGlnLysArgThrGluThr--- 1920
 QY 5428 ATGATATCGGGTACAAACGTGTAGGATGATGAAATAGTGAAGAGT 5487
 Db 1921 -----AsnIleLysAsn-----ArgPhe 1926

QY 5488 GATGAAGATCTAAAAAATGCTGTTCTGTAACCTTCAAGATGATTAATCTATGGGAAA 5547
 Db 1927 AsnGlnGlnTyr-----AsnGlyAspAspPheGlnLys--- 1938
 QY 5548 AATCGGATGAAGAGTACTCGCATCTTCTTCATGCTTTGGCCGAATGGGTGAAGT 5607
 Db 1939 SerSerPheArgSerPheLeuValAspLeuIleArgGlnIleAla-----AlaThr 1955
 QY 5608 TTTTGAACATTAAGAAAAAGCAATTTGCAAAATTTGGAGGCG-----TGTATGAT 5661
 Db 1956 IleAspLysGlyAsnAsnAsnGlyLeuValLysLeuValLysSerValLysCysAsn--- 1974
 QY 5662 TATACTGTGTCATATATGATGATTAAGAAAAATGACGATCGCGTACACAAAT 5721
 Db 1975 -----CysGlyAsnAsnSerGlnAsnGlyLysGlu----- 1984
 QY 5722 AAAAAATTTATAGTACGTGAACACACATGATGAACAAATCAAAATATAGTGAG 5781
 Db 1985 -----GlyGlu 1986
 QY 5782 AATTAAGCAAAATATATTCGACGATCTGTGCAAAAGATGACAGAGCGTCCGAA 5841
 Db 1987 GluAsnAspLeuValLeuLys-----LeuLeuGln 1996
 QY 5842 TATTTAGCAAAACATTAATAAAAAATTTGTCAAAAT-----AAAGTGAATTTGAA 5895
 Db 1997 LysLeuGlnLysLysAlaGlnLysCysLysAspAsnProGluThrSerGlyIleProGln 2016
 QY 5896 TATTAAGTATGAAAGATGTGTCACACACGCGATTAAGTAT----- 5937
 Db 2017 GlnProCys-----GluValSerProAsnIleGlnAspGlnGluGlnProLeuGln 2034
 QY 5938 -----GGTAAATAGTCAAAATATATGCGCCGATCATTAAGCAT-----GAACCA 5979
 Db 2035 GluGlnGluAsnThrValGlnHisProLysIleCysAspAspValLeuLysThrGlnPro 2054
 QY 5980 AAGAGATTGAAGGAAAGTGAATTTGCAAGTGCACAGGTCACACAGTGAACGAG 6039
 Db 2055 GlnProGlnGluProGlnGluThrCysGlnGluLysSerProGlnProHisAspValLysGlu 2074
 QY 6040 GAAACACCGTCACACACGCGGTATCACTGATATCAAAAGCGACGCGATGAAAAAGAACCG 6099
 Db 2075 GluGlnGluGlnGluLys-----LysGlnGlnLysAspLysGlnGln 2089
 QY 6100 AAAACAGCGCGCTTACAAAAACAGCCGAAAAAAGTGAATACTTAACAGAAATGCGA 6159
 Db 2090 ProGlnLeuProProProThrProProAlaIleProSer---SerThrProIle-Pr 2108
 QY 6160 GCACAAACACGACACCGACGACGACGACACAAACACGAAACGACAT 6208
 Db 2108 oProLysProLysProProProGlnValGluLysAsnProTyrProLys 2124

RESULT 9
 T14602
 Variant-specific surface protein - malaria parasite (Plasmodium falciparum) (fragment
 C:Species: Plasmodium falciparum
 C:Date: 20-Sep-1999 #Sequence: Revision 20-Sep-1999 #text_change 09-Jun-2000
 C:Accession: T14602
 R:Yoss, T.S.; Felger, I.; Weiss, N.; Beck, H.P.
 A:Submitted to the EMBL Data Library, February 1998
 A:Description: Identification of a conserved 5' flanking region of Plasmodium falcipa
 A:Reference number: Z18158
 A:Accession: T14602
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2135 <VOS>
 A:Cross-references: EMBL:AF050740; NID:g2944094; PID:g2944095; PIDN:AAC05220.1
 A:Gene: varph17

Alignment Scores: 6.29e-106 Length: 2135
 Pred. No.:

Score:	2303.00	Matches:	681
Percent Similarity:	42.62%	Conservative:	306
Best Local Similarity:	29.40%	Mismatches:	745
Query Match:	11.84%	Indels:	584
DB:	2	Gaps:	95

[illegible]

Dd	310	ProGlnPheLeuAaGtrPrHeGluGluTrrPAlaGluAspPheCysArgLeuTrgYshs	329
Qy	1003	AAATTTGAAAAGGTCACAGGAGCTCCGTGCTG-----AATGACAAAAGAGCTGTAAT	1053
Dd	330	LysIleuLysAspAlaIleGlnLysCysArgGlnLysHisLysGlyAlaLysLysLysLeuTr	349
Qy	1054	TGTAGTCATTAATGACATGATGTGTACGACAACTATTGGAAAAAGATATTTGCATTGG	1113
Dd	350	CysAspPheLeuAaGtrGlyTrAspCysGluGlnThrAlaSerGlyLysHisAspPheGlu	369
Qy	1114	GATATTAAGTGTACTGACTGTTCGACTAAATGCAAACTTTTGAAAGTTTGCTTACGGCAT	1173
Dd	370	GluAspAspCysLysGlyCysGlnTrgSerCysAlaArgPheValAsnTrpIleAspAsn	389
Qy	1174	CACACAGACGATTTTAAAAACAAAAAGAAAAAATATGAAAAAGAAATCAATCATATTTA	1233
Dd	390	GlnLysLysGlnPheLeuLysGlnLysLysTrgGlyThrGlnIleLule---SerLeuLys	408
Qy	1234	TGCAACGATTAACAAATTTGTCAATATTAATTAATGCAATATTAAT-----AAA	1281
Dd	409	SerArgLysLysArgAspAlaGlyLysIleSerThrLysValTrpAspGlyTrcLys	428
Qy	1282	CAATTTTATGAAAACTTAAAGAAAGCAATATGCAACTATGACACTTTTAAATTTTA	1341
Dd	439	LysPheTrgGlnLysLeuLys---SerGluTrgArgThrValGlyLysPheLeuLysLeu	447
Qy	1342	CTAAATGAAGAAAGTATTGT-----AAAGAGGATTCACGAGACAAAAGATATT	1392
Dd	448	LeuAsnAsnGlnLysTrgCysLysGlnValLysAspAspLysGlnLysLysIleAspHe	467
Qy	1393	ACTTTTACTACAGTGTGTGATGACAAAGG-----ATATTTTATGCTTCA	1437
Dd	468	LysThrValAlaSerGlySerAlaSerGlyAspAspValAsnLysThrPheTrgArgThr	487
Qy	1438	GAATTTGCCAAGTGTGTCCGAGCTGGGGGGCTCAAAATGTCAGTGTAAATATCACACAC	1497
Dd	488	LysTrgCysGlnLysCysProTrpCysGlyAlaGlnGlnArgAsnGlyValGlyTrp	507
Qy	1498	AAATAGATTAATGATCGTGAACGCTGAATTAATGAAGACATTAATTAACCTCAGGGGTGTG	1557
Dd	508	LysAlaLysAspAspArgAspCysSerProGlyAsnAspTrgThrLys-----Trg	524
Qy	1558	AAGCCTACTATTAATCACTGCTCTTATAGTGTGAATGAACAAGCTGATATTACACAAAA	1617
Dd	525	LysLysLysGlnIleProIleThrGlyAspLysThrLysSerGlnIleValGlnArg	544
Qy	1618	TTAGAAATTTTGTG-----	1632
Dd	545	TrgArgLysPheCysLysAsnAsnGlyLysAsnGlyAlaAsnGlyArgGlnGlyLysAl	564
Qy	1633	-----AAGCGTCAACCTAATTCACAAAT	1656
Dd	565	GlyGlySerGluAsnGlnLysAlaIleSerAsnSerAspAsnAlaThrThrClyTrgCysCly	584
Qy	1657	AAAAATTAAT-----CAAAATGCGAATGCTATTAT-----AAGCAT	1692
Dd	585	GlyGlnLysSerAspSerSerLeuCysGlnLysTrpThrCysTrgTrgTrgTrgTrgTrg	604
Qy	1693	GAAT-----ATTAATGATGTGAACCTGCAACAAATACCTGAATCAAT	1737
Dd	605	GluAsnAsnAspGlyLysAspIleAsnPheCys---ValGlnGlyAlaTrpGlnAsnSer	623
Qy	1738	AATGATTAATCTAAGATTAATTAATCAATTAATTTTGAATTAATGCGTTCATATTTTA	1797
Dd	624	LysLysAspArgLysValLysSerTrgAsnAlaPhePheTrpAspTrpValHisAspMet	643
Qy	1798	TTAAGAGTACTATTAACTGCAATGACAAACTTAATACTGTATAAAT-----AATACA	1851
Dd	644	LeuIleAspSerIleLysTrpArgAsnGlnHisGlyLysCysIleAsnLysAspAsnGly	663
Qy	1852	ACCACGCAATTGTATGATGAATGATACAGAAATTTGCTTAATGTTTGCACAGATGGGTTAAA	1911

[illegible]

```

QY 3781 -----TGTAAACAGTCGAT 3795
Db 1336 GlnProAlaAlaProThrThrThrProGlyValThrProAlaCysGluValLeu 1355
QY 3796 GATATACCTAAA-----GAAACGATGGAAGAACAGTACAGACTTCATCCAAA 3849
Db 1356 AspleuHeGluLysProLysAsnThrPheLys-----GluAlaCysThrGluLys 1372
QY 3850 ---AAGATATGTAATGATCCGATTCGCAATGC----- 3882
Db 1373 TyrGlyGlyAsnAsnSerArgLeuGlyTrpLysCysIleProThrSerGlyGlyAspLys 1392
QY 3883 -----GGAATATTAATTTAGTGAGACCCCTGCT 3912
Db 1393 AlaAlaThrArgGlySerGlyAspThrThrLysGlnAsnAspSerGluGlySerGluGly 1412
QY 3912 ----- 3912
Db 1413 GluGlySerHisGlnArgAlaLysArgHisThrSerAspAlaSerGlyGluLysSerAla 1432
QY 3913 -----GTGTGATGCCCCCTAGAGACAAAAGTTATGC 3945
Db 1433 LysSerGlyGluProThrGlyGlySerIleCysIleProProArgArgGlyLysLeuTyr 1452
QY 3946 GTACATTTCTGGCAATGATATGAATTAATAAATTAACAATCACAAGTT----- 3996
Db 1453 ValGlyGlyLeuThrLysTrpAlaGluIleGlnSerSerInSerGlnAlaLeuSerGly 1472
QY 3997 -----AATTTAAAGAAAGCTTCATC 4017
Db 1473 GlnThrThrProAlaGlyThrProSerGlnAlaGlnAspProLeuLeuAlaIleVal 1492
QY 4018 AATTCGACACAGACAGAAACATTTCTCATGTAATTAATAA----- 4062
Db 1493 GluSerAlaAlaValGluThrPhePheLeuTrpAspArgTyrLysLysLeuAlaPro 1512
QY 4063 -----AGTAAGATGGTGAAGAAATGAATC----- 4089
Db 1513 GlnSerGlySerSerLeuGlyGlyAlaProLeuGlnLeuLeuAsnGlyAlaIleGly 1532
QY 4090 -----GATAAAGATTAAGAAAGCGAAATTCCTCCCGCATTTTGAGA 4134
Db 1533 SerGluGluThrProGluThrSerLeuLysSerGlyThrIleProProAspPheLeuArg 1552
QY 4135 TCCATGTTCTACACATTTGGAGATATAGAGATTTTATTTGGACA-----GATATA 4188
Db 1553 LeuMetPheThrLeuGlnLysAspTyrArgAspIleCysValGlyValLysGlnAspVal 1572
QY 4189 TCAAAA-----GTCATGTCGAGGAGGATAAA-----CTA 4218
Db 1573 IleLysAlaLeuGlnLysAlaSerGlyAspAsnLysSerSerLysAsnProMetGlnGluIle 1592
QY 4219 AAGAGCAATAGATTCCTTTTCAAAAATGAGGACCAAAAATCTCCT----- 4266
Db 1593 SerSerLysIleGluGluIleLysAsnGlyGlyThrProProThrProValThr 1612
QY 4267 -----AATGAAAAACACGCCAAGAAAGTGGAACAAGACTAGATGAGATA 4314
Db 1613 HisSerProSerSerGlyThrThrProSerSerTrpLysThrAsnGlyGlnHisIle 1632
QY 4315 TGGGAAGCTATGCTATGCTAGCTAGTA-----AAATTTGGCGCAAAAAA----- 4359
Db 1633 TrpLysGlyMetIleCysAlaLeuThrTyrLysGluSerIleGluLysIleGluGln 1652
QY 4360 -----GATGATTTTACGAAACTAGCGTTACAAACGCAATTAATAGTAC- 4407
Db 1653 ValLysAlaThrAspAsnThrAspLeuPheGluLysLeuLysAspLysTyrSerAspTyr 1672
QY 4407 ----- 4407
Db 1673 AspLysValGluIleLysGluGluAsnGluThrGluAlaLysGlyProGlnAspGlyLeu 1692
QY 4408 -----AAAAGCACCATTTTGGAGATTTGCCAAAGCACCCCGATTTTACGATGGCTA 4461

```

```

Db 1693 ThrProGlnThrThrLeuLeuSerAsnPheValLysArgProProThrPheArgTyrLeu 1712
QY 4462 ACCGATCGTACGACGACACTATGCTATACACGCAAAAATATTTAGAAAGTGCGAGAA 4521
Db 1713 GluGluTrpGlyGlnAsnPheCysLysThrArgLysArgMetLeuLysAspIleIleTyr 1732
QY 4522 AATGTAAAGTCAATGACCAATTTGAAGTGTGAT-----ACAGAAATGT 4563
Db 1733 GluCysArgLysSerAspAsnProGlyHisAspTyrCysSerGlyAspGlyTyrHisCys 1752
QY 4564 AATTAAGAA-----TGCAGACACTACGTTAAA 4590
Db 1753 AsnAsnGluLeuLysHisAsnAsnMetPheHisGlyIleAspCysArgAspCysTyrLys 1772
QY 4591 TATATGAAAAAAGAGATGATTCACAAAGATTAATTAATTAACAAAGATGAACGGAC 4650
Db 1773 GluCysArgLysTyrLysLysTrpVal-----HisLysLysPheAspGluPheHisAsnGln 1791
QY 4651 AAAAAAGATTCGATACACACACATTTGTAATGTTACAGAC--TATACGTGA-- 4704
Db 1792 LysAsnLysTyrGlyGluGlnHisGlyLysLeuThrAsnGlyAspAsnTyrSerGlyGly 1811
QY 4705 -----ACGATGCAACAGATTACTTGAACAGAAATTTACTGCTACTGTCGTGAT 4755
Db 1812 GlyAspAsnThrAsnPheCysGlnGlnIleLysGluLysLysThrAla----- 1827
QY 4756 AAGCCTGGAAGTCGCTGTGCTACAAAGAAATATACAAATGTTAGAAAACAGGCTTAC 4815
Db 1828 -----GluAspPhe 1830
QY 4816 TATGATGCCGACAAACATTTGGGTGCACAAAATTTATGAAATGACGCAATATATCT 4875
Db 1831 LeuLysAlaLeuThrArgHisCys-----LysAsnSerGlnAspAspThrAspLysSer 1847
QY 4876 AACATTTCCAGTAAAGTAAG----- 4896
Db 1848 GluGluAspGluLysAsnLysIleAsnPheAspLysProGluAsnThrPheAsnProSer 1867
QY 4897 -----TGCAA-----GATATGTAAGAGCAACACAGCGTATTT 4935
Db 1868 ThrTyrCysLysAlaCysProIleTyrGlyValThr-----CysAsnArgGlyThrArg 1885
QY 4936 AAGTGCAGAAAACAAAGGCTCTAATTAATTAACAATTAAGTGA--GAATTAAGTGA-- 4989
Db 1886 GlyCysArgProLysIleIleAsnGluLysAsnAsnIleGluGluGlnThrAspIle 1905
QY 4990 GATGTGCTTTTCTCTCTGCTGCTACGATATATGTTTTCATGATGATGCGCAATTAAT 5049
Db 1906 AsnIleLeuLeuAsp-----AspGlyAlaThr 1914
QY 5050 ACAGATCCAGAAAGTAAAGATGAAGAAATGGGTGCGAAAAGATTGATGAGTGGCGGCA 5109
Db 1915 AsnAsp-----ThrAspAsnGluLeuHisGluLysCysLysGlu----- 1927
QY 5110 ACGGAAGGTCATTAATTTGGCTCAATTAACAAGAAAAAAGAAAAAGAAAAATTAATAA 5169
Db 1928 -----TyrGlyLeuTyrThrAsnLeuLysGlnGluTrpLysCysGln 1942
QY 5170 ACGTGGATGCGACAAA-----TATTCCTATGAGAGTCCCGCTGTATGCTATGCAATAT 5226
Db 1943 -----HisLysAsnLysSerTyrLysCysGluLeuGlnLysProLeuAsnSer 1958
QY 5227 AGTTTATGATTTTAAGA-----GATATA 5250
Db 1959 GluTyrTyrAspGluArgIleProPheLysIleLeuPheGluArgTyrPheIleAspPhe 1978
QY 5251 ATTCGTAGTATGTAATTTGGAAGATGAAACAAAGAACCGAGGAAATTTGAAGAAA 5310
Db 1979 Ile-----GlnTyrTyrAsnLysSerLysGluArgGlyIleThrArg 1991
QY 5311 ATATTTAACAAAAATGGAACATCAGTTCGCAAGAGAGTGATATGCTACAGGAAATCCC 5370

```


OY	214	GGCTCTTACCCAGTAAATACAGCATATATTATTAATTCACATGTAATTAGAT	273
Db	52	ValGuseSgIySerThAsnGlu-----ThCyAsnLeuVal	64
OY	274	CATAAGAACATACTAATTTTCAGGTATGATGTGAATTGAGACACTCTTCCATGCT	333
		::: :::	
Db	65	GlnAspYrYrYrAsnLysProValTyrGlyAsnSerAsn---ArgYrYrProCysLysAsn	83
OY	334	-----AAGAACAAAACCGATTGTATGATGAGATGAGACAGATCTGAATGTGAAAT	381
		84 LeuYsgIyIleThrAsnGluGluArgPheSerAspRhTrIeuGlyGlnCysThrAsn	103
Db		:::	
OY	382	AAAAATA-----CGAATTTAAAAAGAAAAAT-----GATGCTATAGCTTGTGGC	426
		104 LysLysIleLysGlyAsnGluYrYrSerThLysSerGlyLysAspCysGlyAlaCysAla	123
Db		:::	
OY	427	CCACCTAGAACAGCAGCATATGTGTGATTAACACTGGAGACCTCAATATGATATAAATAC	486
		124 ProYrIArgPheLysIleLysCysSerHisAsnLeuGluSerIleAsp-----ThrThr	141
Db		:::	
OY	487	CAAAATATCATGATTAATTGGGAATGTACTAGTACAGAAATAACGAGGTGTAATCA	546
		142 SerMetThrHisLysLeuLeuLeuGluValAlaCysMetAlaIalaLysTyrGluGlyAsnSer	161
Db		:::	
OY	547	ATTGTATATATATCA-----CATTAAGAAAC-----TCAGACGCTGTGT	588
		162 IleAspRhTrHisTyrProGluHisGlnArgThrAsnGluAspSerProSerGluIleLys	181
Db		:::	
OY	589	ACTGCTCTTGACAGAGATTTGACAGATATAGGTGATATTGTAAAGAGATAGATATGTT	648
		182 ThrMetLeuIalaIyrSerPheIalaSerIleGlyAspIleValAlaGclYrLysAspLeuThr	201
Db		:::	
OY	649	AAACGCAAT-----GTTCATGTAAAGAAAGTAAACCGCTTGTGCAATAGTAA	703
		:::	

Db	202	TYRGlyAsnSerLysGluLysGluLysArgAspGlnLengluThrAsnLeuLysThrIle	221
QY	694	ITTCAGAAAAATCATGATGATGGAATGGAAGATGAA-----GTAAAAATGATTACATCT	747
Db	222	pheGlyLysIleHisGluLysLeuLysAspLysGluGlyAlaGluThrArgTyrGlySer	241
QY	748	GATGGATCTGGAAATTATTATTAATTAAGAGAAGACATGGTGATGATGAATAATAA	807
Db	242	Asp-----ThrThrAsnTyrTyrGlnLeuArgLysAspTrpTyrAlaAsnArgAlaThr	260
QY	808	GTATGGGAAGCATTAACATGATGCATCATTAATATCGGATATTTTATTCACATCA---	864
Db	261	ValTrpGluAlaIleThrCysAspVal-----HisGlySerAspTyrPheArgGlnThrCys	279
QY	865	-----GAAAGTATACACCATATTATTTCAAAATCCTAAATGC-----	900
Db	280	GlyAspLysGluThrThrAlaThrArgValLysAspLysCysArgCysLysAspGluAsn	299
QY	901	GGCCATAACAAGAA-----AAGTCTCTACACATTTAGATTATGTCCTCCAA	948

[illegible]

340 GLULYSLEUGLUGLINCYSARGASPTRYLYSGLNASLLEUTYRCYSSERGLYASNGLY 359

1069 CAGATTTGACGACAACTATTGGAAAAAGGTATTTGCATTGGATAATAAGTACT 1128

220 141ASPcysThrLysThrIleLysLysLysGlyLysLeuValIleGlyGluHisCysThr 379

21 500101CCGACIATTAAGAGTCTTCTGAGTCTTGGTTAGGGAATCAACACAGAAGCATTT 1188
 22 500101CCGACIATTAAGAGTCTTCTGAGTCTTGGTTAGGGAATCAACACAGAAGCATTT 1188

... ..pysnlguseiirpilaspsnnglnlysleugluphe 399

Db 400 LeuLysGlnLysGlnLysTrpGlnLysSerSerLysSerLysGlyLysSer 419
 OY 1237 -----AACGTAACAAATTTGTCATTAATTAATAGTGA----- 1272
 Db 420 GlyGlyValLysGlnLysGlnLysArgGlyValGlyValGlnLysLysAsn 439
 OY 1273 -----TATATTAACAAATTTTGAAGAACTTAAGCAACGCAATATGCACTAAT 1323
 Db 440 TyrAspGlyTrpGlnLysPheTrpLysGlnLysGlnLysSerLysTrpGlyLysVal 459
 OY 1324 GACACTTTTAAATTTACTTAATGAAGAAAGTATGTAAGAGGATTAACAGAGAA 1383
 Db 460 AspAspPheLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 479
 OY 1384 AAGGATATTACTTTTACTTAACAGTCTGAT-----GACAAAGGATATTTTAT 1431
 Db 480 GlnLysLysLysPheTrpLysProLysLysLysLysLysLysLysLysLysLys 499
 OY 1432 CGTTCAAGATTTTCCCAAGTGTCTCCGACTCGGGGCTCAATGTGATTAATAATAC 1491
 Db 500 HisSerGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 519
 OY 1492 ACACACAATTCAGTATGATGTCGTAAGCTGAATGAAGCACTAATAACCTCATG 1551
 Db 520 LysAspLysTrpLysLysLysLysLysLysLysLysLysLysLysLysLysLys 537
 OY 1552 GGTGTACGCTTACTTAATCACTGCTCTTATAGTGAATGAACAGGTGATTAACA 1611
 Db 538 GlyAlaGlnGlyTrpLysLysLysLysLysLysLysLysLysLysLysLysLys 557
 OY 1612 CAAAAATTGAAAAATTTTGTAAAC-----AGCTCAACTAAT----- 1647
 Db 558 ThrLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 577
 OY 1648 -----TACAAGATTAATAATCAAAAA-----TGG 1674
 Db 578 GlyGlyGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 597
 OY 1675 GAATGCTATTATAG----- 1689
 Db 598 LysCysTrpLysGlnLysValGlnLysValLysAspAspLysAsnGlnLysGlnLysGln 617
 OY 1690 -----GATGAAATTAATTAAGA-----TGTAACCTGGAA----- 1719
 Db 618 AspGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 637
 OY 1720 -----CAAAATCTGAAATCAATTAATGATAATCCTAAGATTAATATCAATTCAT 1767
 Db 638 LysHisGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 657
 OY 1768 AATTTTGTGAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1827
 Db 658 AspPhePheTrpLysLysLysLysLysLysLysLysLysLysLysLysLysLys 677
 OY 1828 CTTAAACCTGTTAATTAATAT-----ACAACCAGCATTGATGATGAATGAACAAAT 1884
 Db 678 ValHisSerCysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 697
 OY 1885 TGCTTATGTTTGAAGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1944
 Db 698 CysGlyCysPheLysLysLysLysLysLysLysLysLysLysLysLysLysLys 717
 OY 1945 CTGTTCACAAAAAAAGATATACAGCAATCTATTATTAATTAATTAATTAAT 2004
 Db 718 HisPhe----- 719
 OY 2005 GAAGGTTATTTTAAAGTTATGATTAACCTTGACAAAGATGACAAAGATGAAGAA 2064
 Db 720 -----LysTrpGlnGln 723
 OY 2065 CTTATGAAATATATATAAAGAAAAAATGACTTTTCAATTTTGAATAATTAATAGGAC 2124

Db 724 AlaPheLysAsn-----LysArgGlnLysSerGlyTrpLysPheSerLysLysLysLys 742
 OY 2125 TATTAGAGAAATCAATAGAACTCTTTAGATCACTTAAGAACTGCCAGATATGT 2184
 Db 743 SerAlaAspValValLeuGlnLysLysLysLysLysLysLysLysLysLysLys 752
 OY 2185 AAAGCAATTAATTAACAAACGCAAGCATGTGAACATCCCATATGCAACAAACCCGTGT 2244
 Db 752 ----- 752
 OY 2245 GTTAAACCTGTGAGGACAGCAACCACTAAAAATATAAAGAAATAGCAATCTTT 2304
 Db 753 -----GlnLeuGlnLysLysLysLysLysLysLysLysLysLysLysLys 758
 OY 2305 AAAAGAGTGCATACGAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2364
 Db 758 ----- 758
 OY 2365 CACGAAGGTATTAATTAACGTGGGGGTAGAGAAAGACTTCAAGCAATTTATGTAGA 2424
 Db 759 -----GlnAspLysLysLysLysLysLysLysLysLysLysLysLys 763
 OY 2425 ATATGATTAACATCTTAATCTTAATCTGTTTCAAAATGACCATGTGATGGCAAA 2484
 Db 763 ----- 763
 OY 2485 GGCACAGGTGATGATTAACAAACAAATTTGCTAGCACTGAATGGAATGATCCG 2544
 Db 764 GlyTrpGlyAsp-----ValLysGlnLysLysLysLysLysLysLysLysLysLys 780
 OY 2545 GAACACATGCGTAAAGATCAAGCAAGATGTTATTAATGCTTCCTTAAGAACGACATATATGT 2604
 Db 781 GlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 792
 OY 2605 ACATCAATTTGGAACATTTACAAAGCATGATCACCACCTAATGTAATATGTTGAT 2664
 Db 793 -----ValVal----- 794
 OY 2665 GATTTAGTTAATTAATTCCTTTTGGGGGATGTTCTTATACAAATTAAGAAACAAAC 2724
 Db 794 ----- 794
 OY 2725 AAGATTAATGCAATGTTAATTAAGAAAGATTAACCTTAAGGCCCAAGAAATGACTGAC 2784
 Db 795 -----ValAlaAsp 797
 OY 2785 CAAAACACGACCACTATCTGCGATATACGTTTACAGTTTGCAGATATAGGTGAT 2844
 Db 798 AsnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 811
 OY 2845 ATTAATCGAAGAAAGATCTCTGGAAGAAAGCAAGTGCATGTAAGCTGCAAGACAT 2904
 Db 811 ----- 811
 OY 2905 TTGGAACCTGTTTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2964
 Db 811 ----- 811
 OY 2965 AATGATGATGCCCCAAATTAATTAATTAAGGAAATTTGTTGGGAAAGTATTAAGACC 3024
 Db 812 GlyAspAspAlaAsnAsnLysLysLysLysLysLysLysLysLysLysLysLys 822
 OY 3025 AAAGTATGGGAACCATGAATGATTAATTAATTAATTAATTAATTAATTAATTAAT 3084
 Db 822 ----- 822
 OY 3085 TCACACAAAGTATGTTATTCGGATATAGTATACATACACCAATTTGATTAATCCCA 3144
 Db 822 ----- 822
 OY 3145 CAAAAATTAAGATGATGACCAAGTGGCAAGATGCTGCAAGGTGCAAGAAAGAGAG 3204
 Db 822 ----- 822

QY 3205 TATGATTAAGTGAAGAGAGTGTAGAGATGTAAGATTAAGAT-----AATGCT 3255
 Db 823 -----LysGluLysCysGluLugluThrGlnProLysProProGlyAlaGlyGly 838
 QY 3256 CAAGCTGTAGCAAGAGAGTGTACAGAGTGTACAGAGTGTACAGAGAGTGTATGAA 3315
 Db 839 ProGlyAlaProSerGluThrGlyLugluThrThr----- 850
 QY 3316 TATAATGATTAATAGATTAATGAAGAAACAATGATTAATATCATCAATTAACAA 3375
 Db 851 -----LeuGluAspGluGluGlu 856
 QY 3376 GAATTACATGACACAGCAACAATCTGTGTAATAGTGTATTAAGCTTCCAGTACT 3435
 Db 857 GluLugluAspGluGlu-----GluAspAlaGlyAspGluVal 868
 QY 3436 GCCAAAAATCATATAGACAGAAATGTTATGAATTTTTCGCGAATTAATACCAACAAAT 3495
 Db 868 ----- 868
 QY 3496 GGTGGCAAAAGTAATAAAGTGTACTAGTGAAGAAAGTGTCTGATTCGTTACACAC 3555
 Db 869 -----GluGluGlyLugluThrValAspThr 876
 QY 3556 ACATATGAAATGTTGAGACATATCTCCATGATACAGAAATTTTGATGATTCAGTCA 3615
 Db 876 ----- 876
 QY 3616 CAAATAGATTTTGTGATGAAAAAGTATGATGTAAGATTAACGAAAAATATGCTTTAGA 3675
 Db 877 -----ThrGluGlyAspGluThrGluThrValGluGluInPro 888
 QY 3676 GATTAACACAGACCATGATGTGTGCTGTGTGTGTAAGAGTGAAACCCAGACAG 3735
 Db 889 ValLysAspThrAspArgGlyGly----- 896
 QY 3736 GTACAGATTAATAAGAAAAAGCGAAGAAAGATACGAA-----TGTAA 3786
 Db 897 ---GluGluGluGluLugluLysLysAlaThrAspThrThrThrSerLeuAspValCysAsp 915
 QY 3787 ACAGTAAATGATTAATTAAGAAAAAGATGGAAGAAAGATGATGATTCATCA 3846
 Db 916 ThrValLysAsnAlaLeuThrAsnAsp-----AsnLeuThrAspAlaCysLysLeu 933
 QY 3847 AAAAAGATTAAGTAAGCA-----TATCCGATTCGCAATGC-----GGAAT 3888
 Db 934 LysTyrGlyProGlyLysLysGluArgPheProAsnTrpLysCysValSerSerGlyGlu 953
 3889 ATTAATTTAGTG-----GAAGACCTCGT 3912
 Db 954 LysSerValAlaThrAlaGlySerSerGlyAlaThrGlyLysSerGlyAspLysGlyAla 973
 QY 3913 GTGTGATGCCCTAGAGCAAGAAAGTATTCGTAATTTTGGCAAAATGATTAAGA 3972
 Db 974 IleCysValProProArgTrgArgTrgLeuTrgValGlyGlyLeuThrLys----- 990
 QY 3973 ATTAATAAATTAACAATCAACAATTAATTAAGAAGCTTCAAAATTCGACAGACA 4032
 Db 991 -----LeuThrSerAlaGlyThr 996
 QY 4033 GAACATTTCTTCTCATGTATTAATAAAGTAAGATGCTGTAAGAAATGAACTGCAT 4092
 Db 997 SerSer-----GluSerProGlnGly-----GlySerGluSerSer 1008
 QY 4093 AAAGATTAATAAAGAGCAAAATTCCTCCGATTTTGAATCATCATGTTCTACACATTT 4152
 Db 1009 Arg----- 1009
 QY 4153 GGAGATTATAGAGATTTTATTTGGAACAGATATATCAAAAGTCAAGTGTAGAGGAAGT 4212
 Db 1010 -----AlaSerAspValSerGlnGlyAsnGly----- 1018

QY 4213 AAACATAAAGAGCAAAATAGATTCTTTTCAAAAAATGTGACCAAAAAATCTCTAATGGA 4272
 Db 1018 ----- 1018
 QY 4273 AAACACGCCAAGAAATGTGAGACAGAAACATAGTCATGAGATATGGAGACTATGCTATGT 4332
 Db 1018 ----- 1018
 QY 4333 GCACCTAGTAAATTTGGCGCAAAAAAGATGATTTTACGAAAACTACGTTACAAACAAC 4392
 Db 1018 ----- 1018
 QY 4393 GTCAATTTAGTCAAAAGAACCACTTTGAGGAATTTGCCAAACGCCAGTTTGA 4452
 Db 1019 -----GlyAspAspIleThrThrGluSerLeuArgLys----- 1030
 QY 4453 CGATGGCTAACCGAATGTAAGCAGCAGCTATTCCTATACACAGCAAAAAATTTGAAGAT 4512
 Db 1031 ---TrpPheIleGlu----- 1034
 QY 4513 GTGCAGAAAAATGTAAGTCAATGCAATGCAATTTGAAGTGTATGAAAGAA 4572
 Db 1035 -----ThraIaIa 1037
 QY 4573 TCGAGGACTACGTTAAATATGAAAAAAGAGTGGATTCACAGATTAATAT 4632
 Db 1038 IleGluThrPhePheLeuThrPheIleAspTrgTrgLysGluThrAlaGluLysLys--- 1056
 QY 4633 TACAAGATGAACCGACAAAAAAGATTCGATGACAAACATTTGTTAGTGTACA 4692
 Db 1056 ----- 1056
 QY 4693 GACTATACTGAAACGATGCAACAGATTACTTGAACAGAAATTTACTGTTGTGCT 4752
 Db 1056 ----- 1056
 QY 4753 GATTAAGCTGGAAGTCCCTGTGTGACAAAGAAATATACATTTGTAGAAAAACAGCT 4812
 Db 1057 -----AlaGluLeuGlnArgLysAsnGlyLeuLeu----- 1066
 QY 4813 TACTATGATGCGCAAAACATTTGTGGTGCAAAAAATTTATGAAATGACGCAAAATAT 4872
 Db 1066 ----- 1066
 QY 4873 ACTAATTTGAGTAAGTAAGTAAGTCAAAAGATTAAGTAAGAGCAAAACAGAGTGT 4932
 Db 1067 -----GlyThrGlyAla 1070
 QY 4933 ATTAAGTGGCAAAACAAAGTCTTAATTAATTAATTAATTAAGTAAGTAAGTGAAGAT 4992
 Db 1071 Ser-----LeuAsnLeuGlyGlyAspAspSerAsnProGlnThrGlnLeuGlnLysSer 1088
 QY 4993 GTGCTTTTCTTCTGCGAGCTACGTATAGTTTTCATGATGATGATGCAATTAACA 5052
 Db 1089 GlyThrIleProLeuAspPheLeuArgLeuPheThrLeu-----GlyAspTyrArg 1107
 QY 5053 GATCCAGAAAGTAA-----GATGAAGAAATGGTTGCGAAAAAGATTAAGTGAAGAA 5100
 Db 1108 AspIleLeuValArgGlyValAlaAspAspLysAsnGly----- 1120
 QY 5101 GTGCGGCAACGAAAGGTCATTTGGTCATTAATTAATTAATTAATTAATTAATTAATTAAG 5160
 Db 1121 -----GlyAsnAsn----- 1123
 QY 5161 AAATAAATAAGTCGATGCGCAAAATATTTATGAGGTCCCGCTTGATGCTAG 5220
 Db 1123 ----- 1123
 QY 5221 AAATATGTTTTTATGATTTAAGAGATTAATTTCTAGATTAATTAATTTGAAGATGA 5280
 Db 1124 -----IleIleLeuAsnAlaSerGlyAsnLysAspGlu 1134
 QY 5281 AAACAAAAAGACCGAG-----GAAATTTGAAGAAAAATATTTAACAATAATGGAACA 5331

```

|||||  |||  |||  ::::::::::|  ::::|
Db 1135 LysGlnIysMetGlnIysIleGlnGlnIysIleIleGlnIleLeuProThrSerGlyAsn 1154
QY 5332 TCAGTTGGCAAGAGCATGATAGTACTACAGAAATCCCGTAGTACTGCGCAAAATTT 5391
Db 1155 LysGlnIleThrIleGlyProGlnAAsenSerValAsnAsp 1169
QY 5392 TTCTGGACAGCAAAATTAAGAAATGTGTGGACAGCAATGATATGCGGG 5445
Db 1170 LeuIlePaspArgIleAlaGlnIleHisValIleTrpHisGlyMetValCysAlaLeuThrTyrLys 1189
QY 5446 CGTGTAGGATGATGGAATATAGCAAAATAGTCCAAAGAAAGTGTGAAGATCTAAAAA 5505
Db 1190 - - - - -AspAspAspAsnGly- - - - - 1194
QY 5506 TGTGTCTGTACCTTCAGATGATGATTATCTATGGGCAAAATCCGATGAGATGACT 5565
Db 1194 - - - - - 1194
QY 5566 GCCTATCATGTTTCTCGATGCTTTGCCGAAATGGGGTAGATTTTTCGCAACATAAAGAA 5625
Db 1194 - - - - - 1194
QY 5626 AAGCAATTGAGAAATTTGTTAGGGGCGGTATGATTATTACTTGTGTGATTAATGAAGAT 5685
Db 1194 - - - - - 1194
QY 5686 AAAAGAAAGAAATGTACAGATGCGGTACACAAATATATAAAATTTATAGTAGAGGAAA 5745
Db 1194 - - - - - 1194
QY 5746 CCACAGTATGAAAAACAATCAAAAAAATATGTGAGATTAAGACAAAATATATTCGAG 5805
Db 1194 - - - - - 1194
QY 5806 CATCTGTGGCAAAAGATGAGAGAGAGCGCTCGGCAATATTTAGCAAAACAATTAATAA 5865
Db 1194 - - - - - 1194
QY 5866 ATTGTGAAAATATAAGTGAAGATTTCTGAATATATAGTATGAAGATGTGTCCACACAG 5925
Db 1195 - - - - -LeuLysGlyVal- - - - - 1198
QY 5926 CGATTAACTAGTGTATAGTCAAAATATATGCCGCAATCATTAAGACATGAACCAAAAGAA 5985
Db 1198 - - - - - 1198
QY 5986 GTTGAAGGAAAGTGTATTTGTCAAGTCCCAAGGTCACACAGCTGTACGAAGGAAACA 6045
Db 1198 - - - - - 1198
QY 6046 CCGTCACACAGGCTATCTGATATCAAAAGCGACGCAATCGAAAAAGAAAGCGAAAAACA 6105
Db 1198 - - - - - 1198
QY 6106 GCCGCCCTTACAAACGCCGGAAGAAATGTGAAATCTAACACAGAAATGCCAGCAACA 6165
Db 1199 - - - - -ValLysLysProGlnIleHisValIleGlnAsn- - - - - 1207
QY 6166 ACACGAACCCGACGACAGACACAAACAAACAGAAAGAACATCAACAGAACAAACAACA 6225
Db 1207 - - - - - 1207
QY 6226 GAATCTGACGTGGGACAAATGTTAAAGCCATTCTTTCGATTAACAGATAGCAGGGGT 6285
Db 1207 - - - - - 1207
QY 6286 GGAATAGAGGTTGTATCCAAAAAGATATGACAAATATCTTAATGGGTTGTATTGTA 6345
Db 1207 - - - - - 1207
QY 6346 GGTAAAGTCTAAAGAAATGAAATGCAATATGATGCTCTAGAGAAAAAATTTATGT 6405

```

```

Db 1208 - - - - -ProGlnLys- - - - - 1210
QY 6406 ATAAATATATACATATTTAAATTAAGAACTGAAATAAGCGTCGAATGATATATAA 6465
Db 1210 - - - - - 1210
QY 6466 GAGCTTTTATTAATGTGCAGCAATAGAAACATCAATTTTGTGTGTTAAATATATATT 6535
Db 1211 - - - - -LeuTrp- - - - - 1212
QY 6526 GAAATCTCGACAGCAAAATGCAATTCGAAATGGAACAATTCGACATGATTTAAAGA 6585
Db 1212 - - - - - 1212
QY 6586 ATAAATGATTAATACATATATGATTAAGATATATGTTTTTGGAACTGATATTTCTAAT 6645
Db 1212 - - - - - 1212
QY 6646 GATAAAAAATTTATTAAGTATGTAACAAATAGTGTACAAACATTTCTCAATGAAATTAATAG 6705
Db 1213 - - - - -AsnGlnThrThrLys 1217
QY 6706 AAAAAGAGTAAATAAAAGATGAGAAATTAAGTAATTAATTTTGGGAAAAAATTA 6765
Db 1218 LysProLysAspGlnLysTyrGlnTyrGlnThrAlaLysLeu- - - - - 1231
QY 6766 AAATTTATTTGGAGAAATGATATATGATTAATTAATCATCTACAGACAGAAACGAA 6825
Db 1232 - - - - -GluAspLys 1235
QY 6826 AAAGAAAAATTTAGATTAATTTACCACTACATGACACCAATGAGCCCTCCCTT 6885
Db 1236 GlyGlnLysArgProAspSerSerAlaSerGly- - - - -ThrLysLeuThr- - - - - 1250
QY 6886 GAAGAGTTTGTAAAGAGCCCAATTTTGGAGAGTGTACAGAAATTAATTTTGGGAGAAAGAAATTT 6945
Db 1251 - - - - -AspHeiIleLysArgProPoriLysPheArgTyrLeuGlnIleGlnLysPhe 1269
QY 6946 TGTATTAAGAGGAAGAACAGTGTGTAAATTTGGAGCGCGGCTGTAAAGATATAGTGT 7005
Db 1270 CysLysLysArgThrGlnMetLeuGlyLysIleLysGlnLysPhe- - - - -TyrLysAsn 1287
QY 7006 AATGTGATTAATGACGTTAGACACAAAGATGTCCAGAG- - - - - 7044
Db 1288 GlyLysArgCysSerGlyLysPheLysCysAsnGlnIleValIleAspLysGlnLys 1307
QY 7045 - - - - -GCGTGTACATATCAAAATTTTATTAAGAGGCG 7080
Db 1308 IlePheGlnLysPheLeuLysProThrCysAlaArgHisCysArgPheTyrLysLysTrp 1327
QY 7081 AAACTGAATATGAAGACAAAGAAAGTTCAAAAAGATAAAGATGGCAAAAGTAT 7140
Db 1327 - - - - - 1327
QY 7141 AAGATTAATCTTCTACTGAAAGACATAGAGAGCAACATGTGCTCATGAATATTTA 7200
Db 1328 - - - - -IleAsnThrLysArgAsp- - - - - 1333
QY 7201 AACATGAATTTAAAGAAATTTGTGGCAATTAAGGATTTGTTGTATGCAAAAAACCTCT 7260
Db 1333 - - - - - 1333
QY 7261 TCACAACTACCAAAACAAACAAACAAATCAACATCATCCGATGCTAATGATATGCCAGAA 7320
Db 1333 - - - - - 1333
QY 7321 TCAGTGAATTTATGTTCTGGAAGATTTAACAAGTGTAGTGTCTGAACTTTCAAAAAAG 7380
Db 1334 - - - - -GluPheAsnLys- - - - - 1337
QY 7381 GGATCTATGATTCATACAAAAAATTAAGTAACCTTAATATCCTATGATTTGTAGAG 7440
Db 1337 - - - - - 1337

```

QY 7441 AAAGCAGCATATATTATTCTAAGAGCAAGAAATATATGATATACCTTTGAAGGAA 7500
Db 1338 -----GlnSerAsnAlaIleTyrSerGluInlYsLys 1347
QY 7501 AATTTATACCTATGCTACAAAGAAAGAAAGAAAGTAAATAGTGGACATATAT 7560
Db 1348 LysTyr-----GluGluGluAsnAspSerAlaGluInlYsAsnAsn 1360
QY 7561 AATCTTGGCAGCTTAAGAAACCTTATGCACCTGATTAATATATAGGAAGAA-----AGA 7614
Db 1361 GlYValCysGlyThrLeuLysAspSerAlaIleGlu-----PheLeuAsnThrLeuLysAsn 1379
QY 7615 AACCCCTGGAAATAGAGAAAGAAATGCTTTAAGTAGATTTAATGAATGGAATGTTAC 7674
Db 1380 GlyProCysLysAsnGluSerGluInlYsAsnLysLysAlaGluAspGluIleAspPheLys 1399
QY 7675 AAAAATTCAAAGCTCTATCAGAGCAAAAAAGATGATGTGACCTCCCAAGAAAGAACAT 7734
Db 1400 LysProAspPThrPheLysAspAlaAspAsn-----CysLysPro----- 1413
QY 7735 ATGTCTTAAAGAAATTTAGATGAATTAATTAATGAAGACCTTAAGATAGTAATATATCTC 7794
Db 1414 -----CysSerGluPheLysIleLys----- 1420
QY 7795 CTAAAAATGTTGCTGCAACTGCAAGAAATAGAGAAATAGCATATATATATATATATAT 7854
Db 1421 ----- 7854
QY 7855 TCAGAGAACGGGTGCGCAATGAATCCATATATGAT-----ACTATGAATATATGTTGCTGCT 7911
Db 1426 CysSerSerGlyGlyAsnThrGluGlyLysCysAspGlyLysThrIleAlaIleAlaThr 1445
QY 7912 GATCTGGGTGATGATAGTAGAGAACAA-----GATATGTTAGCAATTTGGTGGTAC 7962
Db 1446 GluIleGluAsnIleLysThrAsnThrLysGluValIleThrMetLeuValIleSerAspSer 1465
QY 7963 TTACCTCCCGTAGAAATTAATATATATATAGTTTGAATACATATATGAAATATGAGAG 8022
Db 1466 LysSerAlaThrGluPheLys----- 1472
QY 8023 AATTAATTAAGGTGAATTAAT 8082
Db 1473 -----AspIleLeuSerGluCysLysAspLysGlyIlePheLysGlyIle----- 1487
QY 8083 GATGCTAATAGAAAGATATTTGGAAAGCAATGACGTGCAGAACCCAGAAAGATGCAAAA 8142
Db 1488 -----ArgLysAspGluThrPgluLys----- 1494
8143 CTTTTAGAAAAGAGATGATGATTTGAACGCATATACATTA-----ATACAGATATAG 8199
Db 1495 -----GlyLysValCysGlyValIleAspIleCysAsnLeuLysLysAspAsn 1510
QY 8200 TGTGACATTAAGAGCAGTACCTGTTGATGATATATATATATATATATATATATATATAT 8259
Db 1511 IleGlyLysGluSerAsp-----LysLysTyrIleIleMetLys-----GluLeuLeu 1526
QY 8260 ACTGAATGCTCTGAAT 8319
Db 1527 LysArgTyrPleuGluLysPhe-----LeuGluAspTyrAsnLysIleLysHis 1542
QY 8320 TCATGATCAGCTGTAATAAATCATCTGACAGATGCCAAGATGATTATGATGAATAATAGTGT 8379
Db 1543 LysIleSerHisCysThrLysAsnGlyLys-----GlySerLysCys 1556
QY 8380 GAACAGCTGTAACAGATGTCAGAAATATATAAATTTGTTCTTAATGGAATCTCTTA 8439
Db 1557 -----IleLysGlyCysValAspLysTyr----- 1564
QY 8440 TTGATATACATCAATAATATACAAAGATTTGATGAACAACAAATATATACAAAAATC 8499
Db 1564 ----- 1564

QY 8500 TCTACTATATGATCATGTTCAAAATTTTGTACAAAAGTTGAAAACCTTTAAGTGAATGT 8559
Db 1565 -----ValGlnGlnLysLysGluGluInlYrPysGlnIleLys----- 1576
QY 8560 TCTGTGAGAGCTTTCTTGTGAATATCTTCATGAAACAAAGTAAGTGTGATATATAATTT 8619
Db 1577 -----GluArgPheAsnGlu-----GlnTyrLysSer 1585
QY 8620 AATGAATAATGAGTCTTCTCAATATATACAAATATATGATCTTTGAGAAACCAAAAAT 8679
Db 1586 LysThrSerAspGluTyrPheAsnValLysSerPheLeuGluThrIleProLys----- 1604
QY 8680 TATTAAGAACCTTGACAGTGTACACTACCTTCTTAAGAAATCATGTAATGTCTTAC 8739
Db 1604 ----- 1604
QY 8740 GATCAAAACAAAGATGATGATAGTAATTAACAATTTTACCTTCTGTCGCAAGAAATGT 8799
Db 1605 -----IleAlaValAlaAsnAsp 1610
QY 8800 TATGATTAATATCTTGTATTAATTTGAAGCATACCTGTCTTAAATAGTTACAGATGATAC 8859
Db 1611 GluAspAsn-----ValIleLysLeuSerLysPheGlyAsnSer 1623
QY 8860 AAAGGTGATGATCTCTCCCAAGAAAGACATTTATGTACAGACCTATACCTGATAT 8919
Db 1624 CysGly-----CysSerAlaSerAlaIleSerThr 1633
QY 8920 AATTTATAGAAAGGTGATTAAGAAATTTTAAAAAAAACTTCTTACCTTGTGCTTACGT 8979
Db 1634 Asn-----GlyAsnGluGlu----- 1638
QY 8980 CAAGGCAATTTGTAGTCAAAATAATTAATTCGAGAGAGAGTGTGCTTTGAGCAATG 9039
Db 1638 ----- 1638
QY 9040 AATATAGTTAGCAGATTAATTCGATATATTAAGAAAGTGAATGATGACACATCTCA 9099
Db 1639 -----AspAlaIleAspCysMet 1644
QY 9100 TTATCTGAATAAATTAATAAATAATTTGAACATCAATGAACACGAAATCGTAAA 9159
Db 1645 IleLysLysLeuGluLysLysIleAspGluLysArgLysProGlyGluAsn----- 1662
QY 9160 ACATGTTGGAAAAATTAATAGAGCTCAGATATGCAACCTATGTATGTATATAATTT 9219
Db 1662 ----- 1662
QY 9220 GCTACTTCAAAAGTAACATTAATGATGAAGATGTGTCATTAACCAAGATGAAGAACT 9279
Db 1662 ----- 1662
QY 9280 AATCAGTTTCTTGTGTTAATTAATGAGCAAGCAACATGTAAGAAACAACT 9339
Db 1663 -----SerGlyGlnThrCys----- 1667
QY 9340 GTAAGTATTCATTAATAACAAAATGTCTCTTCAACAGAAATATTTTGAAGGTCA 9399
Db 1668 -----Asn 1668
QY 9400 GAATTTATAGACAACCTGATGTCAGAAATGATTAATTAATATATATATATATATATAT 9459
Db 1669 GluThrLeuThrHisPro----- 1674
QY 9460 TTGATTAATAATCAATGGAATAATCTAATAATATATATAGCAATTAATAAGATCATCT 9519
Db 1674 ----- 1674
QY 9520 TCAGTAATATAGACAATAAACCATCTGAAGAAATGTTCAGTCAATATATAATCAAAA 9579
Db 1674 ----- 1674
QY 9580 GATTCATCAATGCGCTTTGGAGTTAATGATATATAATGAATAATGTTCAGAGAACAAAAT 9639


```

QY 577 -----TAGAGCTGTACTCTCTGACGAGAGTTTGCAGATATAGTATATGTA 630
Db 192 SerAlaSerGlnLeuGlnThrValLeuAlaArgSerPheAlaAspIleGlyAspIleVal 211
QY 631 AGAGAAATGATATGTTT-----AAACCAATGTCATGACAAAGTA 672
Db 212 ArgGlyLysAspLeuGlnLeuGlyTyrAspAsnLysGlnLysGlnGlnArgLysLeu 231
QY 673 GAAGGGGTCTCCGAGAGTTTTCAGAAATACATGATGAAAGAGAGAGTAA 732
Db 232 GlnGlnLysLeuLysAspIlePheLysLysIleHis-----LysAspValIleLys 248
QY 733 AATGAT-----TACAAATCTGATGGA---TCGTGAATATATATATATTA 774
Db 249 ThrAsnGlyAlaGlnGlnLysArgTyrIleAspAspAlaLysGlnLysAspPheGlnLeu 268
QY 775 AGAGAACGATGTGGATAGTGAATAGAAATTAATGATGAGAGCTATTAACATGATGCA 834
Db 269 ArgGlnAspTrpTrpThrSerAsnArgGlnThrValIleTrpLysAlaLeuIleCysHisAla 288
QY 835 TCATATATATCTGATATTTTATGCAATCAGAAAGTAAATACACCATATTTTCAATCT 894
Db 289 ProLysGlnAlaAsnTyrPheIleLysThrAlaCysAsnValGlyLysGlnThrAsnGly 308
QY 895 AATGCGGCCATTAACAAGAAAGTTCCTACCAATTAATTTGTCCTCAATATTA 954
Db 309 GlnCysHisCysIleGlyLysAspValProThrTyrPheAspTyrValIleProGlnTyrLeu 328
QY 955 CGTTGGTTCAGAGATGGGAGAGAGTTTGGCGAAAGAAATATTAATGAAAG 1014
Db 329 ArgTrpPheGlnGlnLysArgLysAspPheCysArgLysLysLysLysLysLysLys 348
QY 1015 GTCAAGAGCTCTGTGTATGACAAAGACCGTTATTTGTCATTAATGACATGAT 1074
Db 349 LeuGlnLysGlnCysAsnArgAspTyrGlnGlnAsnLeuTyrCysSerGlnLysTyrAsp 368
QY 1075 TGATGCAACATATTTGGAAGAAAGATTTGTCATTTGATTAATAGTACTGACTGT 1134
Db 369 CysThrLysThrIleLysLysGlnLysLysLysLysLysLysLysLysLysLysLys 388
QY 1135 TGACATTAATGCAAGTTTGTGAAGTTGGTATGGAATCAACAGACATTTAAATA 1194
Db 389 SerValTrpPheAspArgMetLysGlnTrpIleAspAsnGlnLysGlnLysGlnLys 408
QY 1195 CAAAGAAATATATGAAAAAGAAATA----- 1221
Db 409 GlnLysArgLysTyrGlnLysIleSerGlyGlyGlySerGlyLysSerProLysArg 428
1222 -----CAATCAATTTATGAAAGATACAAATTTGTCATTAATAT 1263
Db 429 ThrLysArgAlaAlaArgSerSerSerSerSerSerSerSerSerSerSerSer 442
QY 1264 AATAGGATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1323
Db 443 -----GlyTyrGlnLysSerLysLysLysLysLysLysLysLysLysLysLys 460
QY 1324 GACACTTTTAAATTTACTTAATGAGAAAGTATTTGTAAGA-----GGA 1371
Db 461 AspLysPheLeuLysIleLeuAsnLysGlnLysLysGlnLysGlnLysGlnLysGln 480
QY 1372 TTACAGAGGAAAGAGATTTACTTTTACTTAACATGCTGATGACAAAGAGATTTTAT 1431
Db 481 AsnGlnLysAlaAspAsnValAspPheThrAsnGlnLysLysLysLysLysLysLys 499
QY 1432 CGTTGAGATATATGCAAGTGTGCCGACGTGGGGGCAATGATGATGATTAATATAC 1491
Db 500 ArgThrGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 519
QY 1492 ACACCAATATGATATGATGCTGACAGTGAATATATATATATATATATATATATAT 1551
Db 520 LysValLysGlnLysAspLysThrCysGlnSerAlaLysThrLysThrLysAspProLys 538

```

```

QY 1552 GGTGTGAACCTACTAATTAACACTGTCTTATAGTGTGATGACAGAGTATATACA 1611
Db 539 -----AsnIleThrAspIleProValLeuTyrProAspLysSerGlnGlnIleLeu 556
QY 1612 CAAAATTTAGAAAATTTTGTACACACTCACTAATTAACAAAGATATATATATATAT 1671
Db 557 LysLysTyrLysAsnPheCysGlnLysGlnLysAlaPro--GlyGlyGlnLysLysLys 575
QY 1672 TGGGAATGCTATATATAGAT-----GAAATATATATATGATGATTAACAG 1716
Db 576 TrpGlnCysTyrTyrAspArgLysIleHisArgProSerSerLysAsnAsnAsnLysValGln 595
QY 1717 GAACAAAATATCTAATCAATATATGATATCTTAAAGATATATATATATATATATTTT 1776
Db 596 GlyThrTrpAspLysPheThrGlnGlnLysGlnTrpValLysSerTyrAsnValPhePhe 615
QY 1777 GAATTAATGCTTACATATATTAATGAGTATTAAGGATGATGATGATGATGATGAT 1836
Db 616 TrpAspTrpValHisAspMetLeuHisAspSerValGlnTrpLysThrGlnLysSerLys 635
QY 1837 TGATATATATATATACACACG-----CATGTATGATGATGATGATGATGATGAT 1884
Db 636 CysIleAsnAsnAsnThrAsnGlnLysAsnThrCysArgAsnAsnAsnLysCysLysThrAsp 655
QY 1885 TCGTTATGTTTGAAGAGATGGTTAAACAAAAGAAAGAAATGGAATGATATTAAGAA 1944
Db 656 CysGlnCysPheGlnLysTrpValGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGln 675
QY 1945 CTGTTACAAAAAAGAAATATACAGCAATCGTATATATATATATATATATATTTT 2004
Db 676 HisPheGlnLysGlnTrpAspIleValGlnGln-----LysGlyLeuIle 690
QY 2005 GAAGTTATTTTAAAGTATGATTAATTAATGATGATGATGATGATGATGATGAT 2064
Db 691 ValPheSerProTyrGlyValLeuAspLeuVal-----LeuLysGlyLys 706
QY 2065 CTATGGAATATATATATATATATATATATATATATATATATATATATATATATAT 2124
Db 707 LeuLeuGlnAsnIleLysAspValHisGlyAspHisPheLysLysLysLysLysLys 726
QY 2125 TATTTAGAG----- 2142
Db 727 LeuLeuAspGlnGlnAspAlaValAlaValLeuGlnLysAspAsnThrThrIle 746
QY 2143 GAACCTTTGTAGATCACTTAAAGAAAGTCCGACGATGATGATGATGATGATGATGAT 2202
Db 747 AspLysLeuLeuGlnHisGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 764
QY 2203 GAACAGTGTGAACATCCCATATGCAACAAACCCGTGTAAACCTCGTGAGGC 2262
Db 765 GlnLysCysGln----- 768
QY 2263 ACGCAACCACTAATAATATATATATATATATATATATATATATATATATATATAT 2322
Db 769 ----- 773
QY 2323 GAACACGAAATGCTGCTTCAATTAATTAAGAAAGACACAGAGTATATATATAA 2382
Db 774 GluSerArgLysArgSerAlaGln----- 781
QY 2383 CGTGGGGGTGAGAAAGAGCTCAAGGACATTTATGTAATTAATGATTAATAACATTTCT 2442
Db 782 -----ThrArgGlnLysGlnLys----- 787
QY 2443 AATCGTAATCTGTTTTCAAATGACCATGATGATGATGATGATGATGATGATGAT 2502
Db 788 -----ThrGlnGlnProIleAspSerAlaGlyGln----- 797
QY 2503 CAAACAAGATTCTCTAGCACTGAATGGAATGGAATGGAATGGAATGGAATGGAAT 2562
Db 798 -----ValGlnGlnGlnLysAspAspAspAspAspAspAspAspAspAsp 810
QY 2563 CACGAGATGTTATATGCTCTCTAGAAAGACGACATATATATGATCAATTTGGAACAT 2622

```

Db 811 ASPGLUSP----- 813
QY 2623 TTACAAACGGATGATCACCCTTAATGTAAATTTGTGATGATTAGTAAATTC 2682
Db 814 -----
QY 2683 TTTTGGGGGATGTTCTTCTATCAGCAAAATVANGAACAAACAAATATACAAATGTAT 2742
Db 821 GLUGLUGLY----- 823
QY 2743 AAGAAAGAAATAACCTAAAGGGCCCCAAGAGTAACGCCAAACACAGACACT 2802
Db 824 LysGluGluGluThrValThrGluValThrGluValThrGluValThrGluValThr 842
QY 2803 ATCTGTCAGCTATACGTTCACATTTTGCAGATATAGTATATATTCGAGAGAGAT 2862
Db 842 ----- 842
QY 2863 CTCTGGGAAAGAAACGGTCATGTAAAGCTGCAGACATTTGCAACTGTTTTGGT 2922
Db 842 ----- 842
QY 2923 AATATACATAGTCACTCAAGGCAAGCAAAATGATATATATGATGATGCCCCAA 2982
Db 842 ----- 842
QY 2983 TATTTAAATTTAGAGGAAATTTGTGGAGCTAATAGAGCAAAAGTATGGAAAGCATG 3042
Db 842 ----- 842
QY 3043 AAATGATATATAATTTTGAAGATAAATCGGAGCACCAATCAACAAAGTACTTAT 3102
Db 842 ----- 842
QY 3103 TCGGATATATAGTATCATACACATTTGATGATTATATCCACAAAATTAAGATGATG 3162
Db 843 -----Val 843
QY 3163 ACCGAATGGCAGAAATGTACTGCAGAGTGCAGAAAAAGATATAGTAAAGTTGAAAGAG 3222
Db 844 ThrGlu-----GluGlu 847
QY 3223 AATGTAAAGAGTGAAGATAAGATATATGTCAGAGCTGTACGAAAGAGAGTGTACA 3282
Db 848 GlyValLysProCys----- 852
QY 3283 GGTGTACGAAGTGCACAGAAAGCTTGTAAATGATATATATATATATATATATATGAAA 3342
Db 853 -----AspIleValIleGlyLysLeu--- 858
QY 3343 GAACAAATGAAT 3402
Db 858 ----- 858
QY 3403 GTTAGTATATAGTGTATGTAGAGCTTCAGTACTGCCAAAATCATATAGACAGAAATGTT 3462
Db 858 ----- 858
QY 3463 ATTGAATTTTTGTGCAATTTATACAAACAAATGTGTGCAAAAGTAAATAAGTGTACT 3522
Db 858 ----- 858
QY 3523 AGTGAGTAAAGTCTGTATGTGTAATACACACAGTATGAAAATGTTGAGACATATATCTC 3582
Db 858 ----- 858
QY 3583 CATGATACAGAAATTTTGTATGATGTGTCACTCAACAAATGAGTTTGTATGAAAAAAGT 3642
Db 859 -----PheGluAspAspLysSerLeuLysGlu 867
QY 3643 GATGTAAAGATTAACGAAATATATGCTTTAGAGATTAACACAGACATGATGTGCG 3702
Db 868 -----Ala 868
QY 3703 TGTGTGTATAAAGTGTGATGAAACCGACAAAGGTACAGATTAATAAAGAAAAAAGCG 3762
Db 869 CysGlyLeuLysTyrGlyPro----- 875
QY 3763 GAAGAAAAAGATACGGAATGTAAAAACAGTAAATATACTTAAAGAAACGATGAAAG 3822
Db 876 -----GlyGlyLys 878
QY 3823 AAACAATAGAAAGTGTATCTCCAAAAAAGATATATATATATATATATATATATATATAT 3882
Db 879 GluLys-----PheProAsnTyrLysCys 886
QY 3883 GGAATATATAATTTAGT-----GAGACCTCGTGTGTATGCC 3924
Db 887 ValThrProSerGlyValSerThrAlaThrSerGlyLysAspGlyAlaIleCysValPro 906
QY 3925 CTTAGAACACAAAGTTATGCTACATTTCTTGSCA-----AATGATATAT 3969
Db 907 ProArgArgArgArgLeuTyrValGlyGlyLeuSerGlnTyrPalaSerArgGlyLysP 926
QY 3970 GAATATAAATAATTAACATCAACAAAGT-----AAT 3999
Db 927 GluThrThrGluValSerSerGluAlaThrSerAlaProSerGlnSerGluLys 946
QY 4000 TTAAGAGAGCTTTCATCAATCTGCAGACAGCAAGAAATTTCTCTCATGTTATATATAT 4059
Db 947 LeuArgThrAlaPheIleGluSerAlaIleIleLysThrPheLeuTyrPheLysTyr 966
QY 4060 AAA-----AGTAAGAGTCT----- 4074
Db 967 LysGluGluLysLysProAlaThrGlnAspGlyLysLeuGlyValSerLeuPro 986
QY 4075 -----GAGGAAATGAGCTCGATTAAGAAATTAAGAA-----GGCAAAATTCCT 4119
Db 987 GluProSerProGlyGluAspProGlnThrGlnLeuGlnGlnThrGlyValIlePro 1006
QY 4120 CCGCATTTTGTAGATCATCTCTTACACATTTTGAGATTAATAGATTTTATTTTGA 4179
Db 1007 ProAspPheLeuArgIleMetPheTyrThrLeuAlaAspLysAspIleLeu----- 1024
QY 4180 ACAGATATATCAAAAGTCACTGTGAGGAAAGTAACTAAAGCAAAATATGATTCCTT 4229
Db 1024 ----- 1024
QY 4240 TTCAAAAATGTGACCAAAAATCTCTAATGAAAAACGCCAAAGATGTGACAGAA 4299
Db 1024 ----- 1024
QY 4300 CATAGTCATGATATGGAAGCTATGCTATGTGCTAGTAAAAATTTGGGCAAAAAA 4359
Db 1024 ----- 1024
QY 4360 GATGATTTTACCGAAACTACGTTTACACACAGCTCAAAATTTAGTGACAAAAGCACACT 4419
Db 1024 ----- 1024
QY 4420 TTGAGGAATTTGCCAAAGACGCCAGTTTATAGATGCTAACCGAATGTAGACAGAC 4479
Db 1024 ----- 1024
QY 4480 TATGCTATACACGCAAAAAATATTTGAGAGTGTGAGGAAAAATGTAAATGATGAC 4539
Db 1024 ----- 1024
QY 4540 CAATGAGTGTGTATACAGATGTAAATAGAAATGCCAGACTACGTTAAATATATGAAA 4599
Db 1024 ----- 1024
QY 4600 AAAAAAAGAGTGTATCCACAAGATTAATTAACAAGATGAAACGCAAAAAAAGA 4659
Db 1024 ----- 1024

Db 868 -----Ala 868
QY 3703 TGTGTGTATAAAGTGTGATGAAACCGACAAAGGTACAGATTAATAAAGAAAAAAGCG 3762
Db 869 CysGlyLeuLysTyrGlyPro----- 875
QY 3763 GAAGAAAAAGATACGGAATGTAAAAACAGTAAATATACTTAAAGAAACGATGAAAG 3822
Db 876 -----GlyGlyLys 878
QY 3823 AAACAATAGAAAGTGTATCTCCAAAAAAGATATATATATATATATATATATATATATAT 3882
Db 879 GluLys-----PheProAsnTyrLysCys 886
QY 3883 GGAATATATAATTTAGT-----GAGACCTCGTGTGTATGCC 3924
Db 887 ValThrProSerGlyValSerThrAlaThrSerGlyLysAspGlyAlaIleCysValPro 906
QY 3925 CTTAGAACACAAAGTTATGCTACATTTCTTGSCA-----AATGATATAT 3969
Db 907 ProArgArgArgArgLeuTyrValGlyGlyLeuSerGlnTyrPalaSerArgGlyLysP 926
QY 3970 GAATATAAATAATTAACATCAACAAAGT-----AAT 3999
Db 927 GluThrThrGluValSerSerGluAlaThrSerAlaProSerGlnSerGluLys 946
QY 4000 TTAAGAGAGCTTTCATCAATCTGCAGACAGCAAGAAATTTCTCTCATGTTATATATAT 4059
Db 947 LeuArgThrAlaPheIleGluSerAlaIleIleLysThrPheLeuTyrPheLysTyr 966
QY 4060 AAA-----AGTAAGAGTCT----- 4074
Db 967 LysGluGluLysLysProAlaThrGlnAspGlyLysLeuGlyValSerLeuPro 986
QY 4075 -----GAGGAAATGAGCTCGATTAAGAAATTAAGAA-----GGCAAAATTCCT 4119
Db 987 GluProSerProGlyGluAspProGlnThrGlnLeuGlnGlnThrGlyValIlePro 1006
QY 4120 CCGCATTTTGTAGATCATCTCTTACACATTTTGAGATTAATAGATTTTATTTTGA 4179
Db 1007 ProAspPheLeuArgIleMetPheTyrThrLeuAlaAspLysAspIleLeu----- 1024
QY 4180 ACAGATATATCAAAAGTCACTGTGAGGAAAGTAACTAAAGCAAAATATGATTCCTT 4229
Db 1024 ----- 1024
QY 4240 TTCAAAAATGTGACCAAAAATCTCTAATGAAAAACGCCAAAGATGTGACAGAA 4299
Db 1024 ----- 1024
QY 4300 CATAGTCATGATATGGAAGCTATGCTATGTGCTAGTAAAAATTTGGGCAAAAAA 4359
Db 1024 ----- 1024
QY 4360 GATGATTTTACCGAAACTACGTTTACACACAGCTCAAAATTTAGTGACAAAAGCACACT 4419
Db 1024 ----- 1024
QY 4420 TTGAGGAATTTGCCAAAGACGCCAGTTTATAGATGCTAACCGAATGTAGACAGAC 4479
Db 1024 ----- 1024
QY 4480 TATGCTATACACGCAAAAAATATTTGAGAGTGTGAGGAAAAATGTAAATGATGAC 4539
Db 1024 ----- 1024
QY 4540 CAATGAGTGTGTATACAGATGTAAATAGAAATGCCAGACTACGTTAAATATATGAAA 4599
Db 1024 ----- 1024
QY 4600 AAAAAAAGAGTGTATCCACAAGATTAATTAACAAGATGAAACGCAAAAAAAGA 4659
Db 1024 ----- 1024

OY 4660 TTGATGACACACACATTTGTGTATGTTACAGACTATCTAGCAACGATGCAACAGAT 4719
Db 1025 -----
OY 4720 TACTTGAACAGGAATTTACTCTAGTTGTGTGTAAAGCTGGAAGTGCCTGTGTGTA 4779
Db 1032 -----
OY 4780 CAAAGAAATATACATTTGTAGAAAACAGGCTTACTATATGTCGCAAAACATTTGTGG 4839
Db 1032 -----
OY 4840 TGCACAAAATTTATGAAAATGACGCAAAATATCTAACATTTGAGTAAAGATAGTGC 4899
Db 1033 -----
OY 4900 AAAGATTAGTAAAGGAGGCAACACAGGTGCTATTAGTGGCAAAACAAAGTCTTAAT 4959
1048 LysAsnIleValIleuGluAlaSer-----
4960 AACTACAACTTGAAGAATTGACTGAAGATGCTTTTCTCTGCTGACACTAGCT 5019
Db 1055 -----
OY 5020 ATATGTTTTCATGATTCGATGCGCAATTATACAGATCCAGAGTTAAAGATGAAAATGGG 5079
Db 1055 -----
OY 5080 TTGCGAAAAGATTTGATGGAAGTGGCGCAAGGAGGTACAAATTTGGCTCACTACTAC 5139
Db 1056 -----
OY 5140 AAAAGAAAAAAGAAAAAGAAAAATTAACCTCGATGCGCAAAATTTCTTATGAG 5199
Db 1060 -----
OY 5200 GTCCCGCTTGTAGTGTATGAATAATAGTTTATGATTAAAGATATATTTAGCT 5259
Db 1066 -----
OY 5260 ATTGATTAATTTGAGAGATGAAAAAAGACGAGAAAAATTTGAGAAAATATTATAC 5319
Db 1067 -----
OY 5320 AAAAATGAAACATCAGTTGGCAAGAGATGATAGTACAGAAATCCCGTAGTACT 5379
Db 1078 -----
5380 GCGCGAAAATTTTCTGGAACGAAATAAGATGTGTGGAAGCAGATGATGCGGG 5439
1094 ProGlnGlnThrTrpIlePheLysAsnIleAlaLysAspIleTrpAsnIleAlaValCysAla 1113
OY 5440 -----
Db 1114 LeuThrTyrLysGlu-----
OY 5494 GATCTAAAAAAATGTGCTCTTACCTTCAGATGATGATTAATCTATGGGAAAAATGCG 5553
Db 1122 -----
OY 5554 GATGAGTACTGCTATCAGTTCTTCGATGTTGCCGATGGGGAAGATTTTTCG 5613
Db 1123 -----
OY 5614 AAACATAAAGAAAGAAATTTGCAAAATTTGATGAGGCGGTGAATTAATCTGTGT 5673
Db 1125 -----
OY 5674 GATTAATGAAGATAAGAAAGAAATGTACAGATGCGGTACACAAATATAAAATTTAT 5733
Db 1125 -----

OY 5734 AGTAGTGAACACACAGTATGAAAAACAATCAAAAAATATGCTGAGATAAGACAAA 5793
Db 1126 -----
OY 5794 ATATATTCGAGATCCTGTGGCAAAAGATGACAGAGAGCGTCCGAATTTTACGAAA 5853
Db 1133 -----
OY 5854 CAATTAAAAAAATTTGTGAAAATAAGTGGAGATTGTGAATATTAAGTAAAGAT 5913
Db 1134 -----
OY 5914 GTGTCCACACAGGATTAATGATGTAATAGTCAAAATATGCCGATCATTAAGAGAT 5973
Db 1137 -----
OY 5974 GAACCAAAAAGTTGAGAAAAGTATATGTCAGTCCACAGAGTCCACACAGTGTGA 6033
Db 1147 -----
OY 6034 CCAAGGAAACACCGTCACACGCGGTATCATGATATCAAAAAGCAGCGCATCGAAAAA 6093
Db 1147 -----
OY 6094 GAAGCGAAAACAGCGCGCTTACAAAACAGCGGAAAAAGTGAATTAACACAGAA 6153
Db 1147 -----
OY 6154 ATGCGAGCAAAACACGAAACCGAGCAGCAGCAACAAACAGAAACGAAATCAACA 6213
Db 1147 -----
OY 6214 GCAACACACAGATCTGACGTGGGCACATGTTAAAGCCATTTCTTGATTAACCA 6273
Db 1147 -----
OY 6274 GATACAGCGGTGAATAGAGGTTGTAATCCAAAACGTATGACAAATATCTAAATG 6333
Db 1147 -----
OY 6334 GGTGTATTGTAGTAACTTAAGAAATAAGAAATGAGCATATGTATGCTCTAGAGA 6393
Db 1147 -----
OY 6394 AAAAATATGATTAATAATATATCAATATTAATTAATGAACTGAAATTAAGCTGAC 6453
Db 1148 -----
OY 6454 AATGATTAAGAGAGGCTTTATTAATGTCACCAATAGAACTCAATTTTGTGTTA 6513
Db 1159 -----
OY 6514 AAATATTAATTTGAAATCTCGACAGAAATTAATGCAAAATGAGCAATTCAGAT 6573
Db 1159 -----
OY 6574 GAATTTAAAGAAATAGTATTAATACATATGTCATTAATAAGATATGTTTTGAACT 6633
Db 1159 -----
OY 6634 GATATTTCTAATGATTAATAAAATTAATAGCTGAACAAATAGTAAACCAATTCGAAT 6693
Db 1160 -----
OY 6694 GAAATATTAAGAAAAAAGAGATTAATAAAAGATGAAGAAATTAAGTAAATTTTGG 6753
Db 1161 GluSerGlyAlaLysSerAsp-----
OY 6754 GAGAAAAAATTAATTAATTTGGGAAGATATATGATTAATTAATCATCATCACA 6813
Db 1168 -----
OY 6814 GACGAAAAGAAAAAAGAAATTTAGAGATTAATTAACAGTAAATGACATGACCAAACTG 6873

[illegible]

QY 8953 AAAAACTTCTACTCTGCTTTCAGTCAGACAAATGTTAGGTCAAAAAATATAATATCG 9012
 Db 1529 -----
 QY 9013 GAAAGAGTTGTCTTGGAGCAATGAATATAGTATGCAATTTTCAGATATAT 9072
 Db 1530 -----
 QY 9073 AAGGAACGTATGATGACACCTCATTTATCGAAAAATTTAAAAAATATTGAACA 9132
 Db 1537 LysProCysaspGlyLeuasp-----ThrgLysPheLysAsnValIle 1536
 QY 9133 TCAATATAGACCAACCAAAATCGTAAACATGCTGGAAAAATATAGACGTAGATG 9192
 Db 1543 -----
 QY 9193 CAGCATATGTTATGTGATATAAATGCTACTTCAAAAGTAACTTATAGTGAAGATG 9252
 Db 1543 -----
 QY 9253 TGTCAATTCACAAAGATGAGAAACATCACTTTCTTCGTTGTTAATTGATGGCA 9312
 Db 1544 -----
 QY 9313 AAGCAAGCATGTAAGAAAAAAGAACATGTAGTATCATTTAAAAACAAATGCTCGT 9372
 Db 1546 -----
 QY 9373 TCAAGCAAGATATTTTGGAGCGTCAGATATTTAAGACAACTGATGTCAGATGAT 9432
 Db 1552 AsnSerThrAspAsn-----SerGlnAsnLysAsnAsnAsp 1564
 QY 9433 ATTAGAAAAATATTAGCTTGAATATATGATATAAATAATCAATGGAATAATATA 9492
 Db 1565 -----
 QY 9493 AATATATCAATTAAGATCAATCTTCAGTAAATATAGACAAATACCATCTGAAGA 9552
 Db 1576 LysIleSerGlnCysLysGlnLysHisSerGlyGlnThr--ProCysasp--- 1593
 QY 9553 AATGTCTACATATTAATAATCAAAAGATCTCATGCGCTTGGAGTAAATGATATA 9612
 Db 1594 -----
 QY 9613 AATGAATATGTATACAGCAACAAAAATATGAATAAT----- 9651
 Db 1612 GlnGlu-----GlnAsnProGlnAsnLysValGlnGlnProLysPheCys 1626
 QY 9652 ---GAATTCAGAAAGTACTATAAAAAATATATCTCGTGTATTTGTTGAAGTGA 9708
 Db 1627 ProAspMetLysGlnProLysLysGlnAsnAspGlnGlnGlyThrCysGlyLysP 1646
 QY 9709 ACACCAAAATCATGTACTAGATGAATATA-----AAGAAAGAGCAACAGTT 9762
 Db 1647 GlnGlnLysLysLysValGlnAspSerValIleGlnGlnLysGlnGlnLysLys 1666
 QY 9763 CGTCTAAAGCACTCATTTCTTACACCCCACTAGANTCTTTTATACACACCTTTA 9822
 Db 1667 AlaProGlnLysProProLysProLysPro-----GlnAlaProLys 1680
 QY 9823 TTCTCAACATCGAGTAGACCAATATGATCT----- 9855
 Db 1681 LysGlnGlnAsnValIleProLysProProProProLysLysArgArgIleLysThr 1700
 QY 9856 AAAAAT-----GATATATGAAGTACTATCTCTGTTATATGATATG 9900
 Db 1701 ArgAsnValIleAspHisProAlaValIleProAlaLeuMetSerSerThrIleMet 1720
 QY 9901 GCGTATAGTTTG-----ATACGCTCATTTCAAGCAAGAAAAATTCAAATCG 9948
 Db 1721 SerIleGlyIleGlyPheAlaLysPheThrTyrrPheTyrrLeuLysLysLysThr 1740

QY 9949 TCTGTG---GACTTGTGGTATCTGAATATCCCGCAAGAGATATGGAATGCCCTAG 10005
 Db 1741 SerValGlyAsnLeuPheGlnIleLeuGlnIleProLysSerAspLysPheProThr 1760
 QY 10006 TTGGAATCCAAAAATAGTACATATACATATAGAGTGGTCCATATTAAGCAAAATAT 10065
 Db 1761 LeuLysSerSerAsnArgLysIleProLysLysAspArgHisLysGlyLysThr 1780
 QY 10066 ATATATATGAGAGACATCTAGTGGAGATGAAGATTAATATATGAGGACATATCTCC 10125
 Db 1781 IleTyrrMetGlnLysAspSerSerGlyAspGlu--LysTyrrAlaPheMetSerAspThr 1799
 QY 10126 TCTGATATCTTCAATCCGAAGTGGATGAAGATTTGGAATTTGATATATATATGTA 10185
 Db 1800 ThrAspIleThrSerSerGlnLysGlnGlnLysLysLysLysLysLysLysLys 1819
 QY 10186 CCAGTATGCTCAATATATAAATCATGATAGAGTACTAGAACATCAAAAGGAT 10245
 Db 1820 ProGlySerProLysLysLysThrLeuIleGlnValIleLeuGlnProSerLysArg 1839
 QY 10246 ATACCAAGTAT-----GATACCAAGTAAATGATACACACGT 10284
 Db 1840 ThrGlnAsnAspIleHisAsnAspIleProSerAspIleProAsnSerAspThrPro 1859
 QY 10285 ACGAATATATTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10344
 Db 1860 Pro-----LleThrAspAspLysLysLysLysLysLysLysLysLysLys 1877
 QY 10345 TTA-----CCAATATACAGACCAAAATATAATTAACAAAGTACAGATATTCAT 10395
 Db 1878 LeuGlnAsnThrGlnAsnThrGlnProAsnIleLeuHisAspAsnValAsp-----Asn 1895
 QY 10396 AATACAGAACCTAATATCTTATATCTGATATATCTGATATATCTGATATATCT 10455
 Db 1896 AsnThrHisPro---ThreSerArgHisAsnMetAspGlnLysProPheIleMetSer 1914
 QY 10456 ATTCATGATAGGATTTATATCTGGAAGAAATATGATAT----- 10497
 Db 1915 IleHisAspArgAsnLeuPheSerGlyGlnLysLysLysLysLysLysLysLys 1934
 QY 10498 -----AATATATATATGATCTAT-----ACTAATAT 10527
 Db 1935 AsnAsnProLysIleAsnIleSerAspSerThrAsnSerMetAspSerLeuThrSerAsn 1954
 QY 10538 GATATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10587
 Db 1955 HisSerProLysAsnAspLysAsnAspLeuTyrrGlyIleAspLeuIleAsnAsp 1974
 QY 10588 CTAGGTGCTAAACCTATTTATATATATGATGATGATGATGATGATGATGATGAT 10628
 Db 1975 Leu--SerGlyAsnHisIleAspIleTyrrAspGlnMetLeu 1987
 RESULT 12
 T14029
 variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
 N:Alternate names: erythrocyte membrane binding protein 1 (EMPI)
 C:Species: Plasmodium falciparum
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
 C:Accession: T14029
 R:Chen, Q.; Barrigan, A.; Fernandez, V.; Sundstrom, A.; Schlichtherle, M.; Sahlen, A.
 J. Exp. Med. 187, 15-23, 1998
 A:Title: Identification of Plasmodium falciparum erythrocyte membrane protein 1 (PfEM
 A:Accession: T14029
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2228 <CH>
 A:Cross-references: EMBL:AF003473; NID:92961467; PID:92961468; PIDN:AAC05730.1
 C:Genetics:
 A:Note: PCR351.2-var1
 Alignment Scores: 7.61e-78 Length: 2228
 Pred. No.:

Score: 1731.50 Matches: 771
Percent Similarity: 29.10% Conservative: 326
Best Local Similarity: 20.45% Mismatches: 659
Query Match: 8.90% Indels: 2016
DB: 2 Gaps: 134

US-10-087-013-1 (1-10628) x T14029 (1-2228)

OY 97 AAAAGTCGAGAAATGTTTGAACGTTATGCCAAAATTAAGACATCCATCAATAT 156
Db 13 GLUAspAlaIysHisValLeuAspGluPheGlyGlnIysVal---HisAspGluValHis 31
OY 157 -----GCAAAAGACATGCGATTCGTTGAAAAGCGATTCGAGAAAGCAATTCGT 210
Db 32 GlyIuAlaIysAsnIysValSerGluIysGlySerIleuAlaSerIleu 51
OY 211 GGTGTCCTCTACGCGCAATTAAGCAATTAATATATATCATCATCATTAATTA 270
Db 52 GlyIuThrAlaPheThrValIys----- 59
OY 271 GATCATTAAGCAATCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 318
Db 60 -----SerMetGlnThrGluSerIysTyrThrGluLeuIleGluAlaAsnSerIysArg 77
OY 319 CATCTTCG-----CATGTCAGAAACAAACCGATTCGATGAGATGAAGA 366
Db 78 AsnProCysIysAspGlyIysGlyAsnAspValAspPheSerValIysGln 97
OY 367 TCGATATGTCGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 423
Db 98 AlaGlyTyrAspAsnIys-----LysMetIysCysSerAsnGlyMetThrCys 113
OY 424 GCGCAGACATTAAGCAATCATTAATTAATTAATTAATTAATTAATTAATTAAT 483
Db 114 AlaProPheArgIysLeuHisLeuCysAsnIysAsnProAsnMetAsnSerAsnAsp 133
OY 484 ACCCAAAATAT---CATGTTTATGCGAAATGCTACTACAGCAAAATACGAAGT 540
Db 134 SerSerIysAlaIysHisAspLeuAlaGluValCysMetAlaAlaIysIysGlnGly 153
OY 541 GAATCAATGCTTAATTAATCAACA-----CATTAAGCAATTCGA---GAC 582
Db 154 GluSerIleIysThrHisIysTyrProIysTyrAspSerIysTyrProGlySerAspPhePro 173
OY 583 GCTTGTACTGCTCTGCACGAAATTTGCAATATAGATATTTGTAAGAGAAATAG 642
Db 174 MetCysThrMetLeuAlaArgSerPheAlaAspIleGlyAspIleIleArgGlyArgAsp 193
OY 643 ATGTTT-----AAACCAATGTCATGCAAAAGTA 672
Db 194 LeuTyrLeuGlyAsnIysIysIysGlnAsnGlyIysGluThrGluArgGlnIysLeu 213
OY 673 GAACGCGCTCCGAGAGGTTTCAAGAAATATACATGATGAATGAAGAT---GAAGTA 729
Db 214 GluGlnIysLeuIysGluIlePheIysIleHisAspAsnIysAspIysGlnAla 233
OY 730 AAAAATGATTAACATCTGATGATCTGAAATTAATTAATTAATTAAGAGATGCTGT 789
Db 234 GlnIysArgIysAsnGlyAspGluAspProAsnPheTyrIysLeuArgGlnIysAsp 253
OY 790 AATGTGAATGAATAAGATATGGAAGCTATTAACATGATGATCATATAT---AAATCT 846
Db 254 ThrAlaAsnArgIuThrValIleTyrGlyAlaMetThrCysSerIysGluLeuAspAsnSer 273
OY 847 GATATATTTATGCAATCAGAAATTAATCA-----CCATTAATTTCAAAATCCTAA 897
Db 274 SerTyrPheArgIleThrCysAsnAspThrGlyGlnGlyProSerGlnThrHisAsnIys 293
OY 898 TGC-----GCCATTAACAAAGAAAG----- 918
Db 294 CysArgCysAspIysAspIysGlyAlaAsnAlaGlyIysProIysAlaGlyAsp 313
OY 919 -----GTTCCTACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 969

Db 314 ValThrIleValProThrTyrPheAspTyrValProGlnTyrLeuArgIlePheGlnGlu 333
OY 970 TGGGAGAAAGATTTTCCGAAAAGAAATATTAATTAATTAAGTCAAGGACTCTCTGT 1029
Db 334 TrpAlaGluAspPheCysArgIysIysIysIysLeuGluAsnGluIysGlnIys 353
OY 1030 CGT---AATGACAA-----GAACGCTTATATGATGATTAATGACATGATTTGAC 1080
Db 354 ArgGlyIysAspIysSerAspGluTyrArgTyrCysSerArgAsnGlyTyrAspCysGlu 373
OY 1081 ACAACTATTGGAAGAAAGATTTTGCATTTGATTAATTAATTAATTAATTAATTAAT 1140
Db 374 GlnThrIleSerArgIysGlyIysValArgMetGlyIysGlyCysThrAspCysPhePhe 393
OY 1141 AAATGCAAAAGTTTGAAGTTGTTGAGATTAACAAACAAACATTAATTAATTAATTA 1200
Db 394 AlaCysHisSerIysGluAsnThrPheAspAsnGlnArgIysGlnPheAspIysGlnIys 413
OY 1201 GAAAAATATGAAAAAGAAATACATCATTAATTAATGACATTAACAAATTTGCAATTA 1260
Db 414 ---LysTyrThrIysGluIleSerAspGlyGlyIysArgIysIysArgAlaValGlyGly 432
OY 1261 ATTAAT-----AGTCATATTATTAACAAATTTATGAAAACTTAAGAAAGCAATAT 1314
Db 433 ThrThrIysTyrGlnGlyIysTyrGlnIysSerPheTyrGlnIysLeuIysAsnAspGlyTyr 452
OY 1315 GCAACTAATGCACTTTTAAATTTACTTAATTAAGAAAGAAATTTGTAATTAATTA 1365
Db 453 GlyThrValAspAlaPheLeuGlyLeuLeuAsnGlnIysAlaCysIysAspIleThr 472
OY 1366 -----GAGAGA 1371
Db 473 AspGlyIysIysIleAsnPheIysGluValAlaAsnSerGlyGlyValIleGlyGly 492
OY 1372 TTACAGGAGAAAAAGATATTACTTTACTTAACAGTCTGATGACAAAGGATTTTAT 1431
Db 493 SerGlyIysThrSerGlyAlaSerGlyThrAsnAsp---GluAsnIysGlyThrPheTyr 511
OY 1432 CGTTGAGAAATATTTGCCAAGTGTGCCGACTGCCGGGTCAAAATGATGATTAATTAAT 1491
Db 512 ArgSerGluTyrCysGlnProCysProAspCysGlyAlaGln----- 525
OY 1492 ACAACAAATCAGATTAATGATCTGAAACGTGAAT-----AAT 1530
Db 526 ---HisIysGlyIysGlnGlnIleArgIysThrIysValIysIysMetAlaGlyPheSer 544
OY 1531 GAAGCATATTAACCTCATGCGGTGAGACCTTAATTAATCACTGCTTTATAGTGT 1590
Db 545 LysLeuTyrIysProIleAsnGlyIys-MetValIleLeuLeuIysSerIleuIysValIa 564
OY 1591 AATGACAAAGTGTATTAACAAATAATTA---GAATAATTTTGT-----AACAGCTC 1640
Db 564 IlysAspMetMetIleLeuIysIysAsnTyrIysGluPheCysIleuThrGlnAsnSerSe 584
OY 1641 A-----ACTAATTAACAAAGATTAATTAATTAATTAATTAATTAATTAATTA 1671
Db 584 AspGlySerValGlySerValIleThrThrGlyAlaSerGlyGlyIysSerGlnIys 604
OY 1672 -----TGGAATGCTTATTAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1715
Db 604 SGIuLeuTyrAspGluTyrPheCys---TyrIysHisAsnGluValGlnIysValAsnVa 623
OY 1716 GGAACAAATATGCAATCAATTAATGAT----- 1743
Db 623 IeGlnGlyIleValGlnGluIysAspAspGluLeuIysGlyAlaGlyIleIysLeuIysIle 643
OY 1744 ---AATCTTAAG----- 1752
Db 643 uProAsnProIysIysAsnIysGluValIserGluAlaIysSerGlnAsnHisAlaAs 663
OY 1753 ---ATAATATCAATTAATTAATTTTGAATTAATGCTTAACATTAATTAATGAAGGATAC 1808


```
Db 663 pIleGlnIysThrPheHisAspPheTyrTyrValAlaHisMetLeuLysAspSe 683
QY 1809 TATTAAAGTGAATGACAAA---CTTAAAACTGTATTAATAT---ACAACCAGCATGTG 1862
Db 683 rIleHisrPrpGthrLysArgLeuLysSerCysIleSerSpIyLysThrMetLysCy 703
QY 1863 TATGTAGTAATGTAAACAAATTCCTTATGTCTTTTGACAGATGGGTAAACAAAAGAAGA 1922
Db 703 sArgAsnGlyCysAsnLysLysCysAspCysPheGluLysrPrpValLysGlnLysGluTh 723
QY 1923 AGAATGGAATAGTAAAGAAACCTGCACAAAAAAAAGAAATATACAGCATGTCTTA 1982
Db 723 rGluTrpLysrProIleLysAspHisPheLysThrGlnGluLysIleProGluGlyTyr 743
QY 1983 TAGTAATATTAAATCTTTTGAAGGTATTTTAAAGTATGATTAACCTTGACAA 2042
Db 743 r----- 743
QY 2043 AGATGAAGCAAAATGAAAGAACTATGAAAAATATAAAAGAAAAAAATGACTTTC 2102
Db 744 ----- 744
QY 2103 CAATTGGAAAAATATAGGACTATTAGAAATGCAATAGAACTCTTGTAGATCACTT 2162
Db 745 rThrLeuGlu-----LeuIleLeuLysLeuGlnPheLe 756
QY 2163 AAAAAGACGCCACGATATGTAAAGCAATTAATACAAACGAAGCATGTGAACATCCCA 2222
Db 756 ulysGlu-----AspThrGluGluAsnThrGluAsnSerLe 768
QY 2223 TAATGCACAAACAAACCCGTGTGTAAACCTGTGAGGACGCCAACCCCTAAATAT 2282
Db 768 uAspAla----- 770
QY 2283 AAAAGAAATAGCAATACTTTTAAAGAGTGCATACGAGGAACGCAAAATCGTGCT 2342
Db 771 ----- 773
QY 2343 TCATAATTGAAAGAAAGGACACAGAGGTATATTAACGTGGGGGTAGAGAAAGCA 2402
Db 773 ----- 773
QY 2403 CTTCAAGACAAATTTATGTAGATTAATGATTAACATCTTAATCTGATCTTTC 2462
Db 773 ----- 773
QY 2463 AAATGACCATGTGATGCAAGGACACAGGTGATGATCAAAACAAAGATTGTGTAGG 2522
Db 773 ----- 773
2523 AACTGATGGAAGTGTATCCGGAACACATGCGTAAAGATCAAGAAATGTTATATGCC 2582
Db 773 ----- 773
QY 2583 TCCTAGAGAGACATATATGTACATCCAAATTTGAAACATTTACAACGATGATCACCC 2642
Db 774 ----- 780
QY 2643 ACTTAATGTAATATTGTTGATGATTTAGTAAATTCCTTTTGGGGAGATGTTCTTCT 2702
Db 780 ----- 780
QY 2703 ATCAGCAAAATATGAGCAAAACAGATATATGCAATGATTAAGAAAGAAATTAACCTTAA 2762
Db 781 ----- 792
QY 2763 GGGCCCCAAAGAACTAGCAACCAACCAACCAATATCTGTGAGCTATATACGTTA 2822
Db 792 aValValAsnAlaGlyThrGlu-----GlnLysThrLeu----- 803
QY 2823 CAGTTTGCAGATATAGGTGATATATATCGAGAAAGAAATCTCTGGGAAAGACGTGA 2882
Db 803 ----- 803
QY 2883 CATGTAAAGCTGCACAAAGACATTTGCAAACTGTTTTGGTAAATATATACATTAAGTCACTCA 2942
Db 804 -MetAspLysLeuLeuAsnHis----- 810
QY 2943 AGCCAAAGAAATGATTAATATATATGATGATGCCCCCAAAATATTTAAATTGAGGAAAA 3002
Db 810 ----- 810
QY 3003 TTGGTGGGAAGCTAATAGAGCCAAAGATATGGAAGCCATGAATGTGATATTAATATT 3062
Db 811 ----- 823
QY 3063 G-----AAGGATTAATCGGGACACCAATCAACACAAAGTACTTATCGCGATATAGTGA 3116
Db 823 uProGluGluAsnSerArgLysrPrpValrGlyArgSer-----AlaAs 835
QY 3117 TCATACACCATTTGATGATTAATCCACAAAATTAAGATGATGACCGAATGGGACAGA 3176
Db 835 pProSerPro---AspIlePheIlePro----- 843
QY 3177 ATGCTCTCAAGGTGCACAAAAAGAGATATGATATAGTTGAAGAGAAAGTATAGAGAGTG 3236
Db 844 ----- 852
QY 3237 TAAAGTAAAGATTAATGTCACAGGCTGTACGAAAGAGAGTGTACAGGTGTACAGAGTG 3296
Db 852 ----- 852
QY 3297 CACAGAAAGCTTAAATGATTAATGATTAATTAATGATTAATGAGAAAGAAATGGAATAT 3356
Db 853 ----- 857
QY 3357 AATATCAGTAAATATCAAAAGATTTACATGACACACAAATGTCTGTAGTAATAGTG 3416
Db 858 ----- 876
QY 3417 TATTAAGCTTCCAGTACTGTCACAAAATCATATAGACAGAAATGTATTAATTTTGTG 3476
Db 876 aThrAspThrThrSerLeu----- 883
QY 3477 GGAATTTATCCACAAATGTGTGCAAAAGTAATTAAGTGTACTGATGAAGTGC 3536
Db 884 ----- 887
QY 3537 TGTCAATGTACTAACACACGATATGAATATGTTGAGCATATCTCCATGATTAAGGAAA 3596
Db 887 olleValGlyLysValLeuThr----- 894
QY 3597 TTTTGATGATTTGTCAGTCACAAAATGAGTTTGTGATGAAAAAAGTATGATGATGAATAA 3656
Db 895 ----- 897
QY 3657 CGAAAAATATGCTTTTAGAGATTAACCAACAGACCATGATGTGCTGTGTGTAAG 3716
Db 897 nGluSerLeu-----GlnAsp-----AlaCysSerLeuLysTy 908
QY 3717 TGAATCGAAACGACAAAGGTATGATTAACCAAAAGGAGGAGAAAGAAAGATAC 3776
Db 908 rGlyGlyAsnAsnSerArgLeuGlyTrpArgCysValThrProSerLysGluProThr 928
QY 3777 GGAATGTAAACATGATGATATATCTTAAAGAAACAGATGAAAGAAACAGTACAGA 3836
Db 928 r----- 928
QY 3837 TTGTATCCAAAAAGATATGATATCCGATTTGGCATGCGGAATATTAATTT 3896
Db 929 ----- 934
QY 3897 AGTGAAGACCTGTGTGTATGCCCTAGAAAGCAAAAGTATGCTACATTTCTT 3956
Db 935 ----- 950
QY 950 AlalIecysValProProArgArgArgLeuThrIleLysLysI 950
```

QY 3957 G-----GCCAATGATATGAATATAAATTT 3963
Db 950 eValaSPTRPalaThrlYstThrGlUserProGlnAlaSerGlySerGluAlaSerSerth 970
QY 3984 A-----CAATCACAAGTTAAATTTAAAGAGCTTCATCAATC 4022
Db 970 rSerGlySerThrThrProPaspSerLysGluAlaLeuLeuLysAlaIleValaIleGluSe 990
QY 4023 TGCAGACGACGAAACATCTCTTCATGCTATATATATAA- -AG 4064
Db 990 rAlaAlaIleGluThrPhePheLeuTrpHisArgTyrLysGluGluLysAlaValAlaI 1010
QY 4065 TAAGATGTGGAAGAAATGAATC----- 4089
Db 1010 agInGluGluAlaGlyHisGlyLeuProArgValGluGluGlySerProGluTyrTrsPpr 1030
QY 4090 -GATAAAGATTTAAAGAGCAAAATTCCTCCGCAATTTTGAGATCCATGTTACAC 4148
Db 1030 ogLAspLysLeuLysGluGlyLysIleProAspGlyPheLeuArgGlnMetPheTyrTh 1050
4149 ATTTGAGATTTATAGAGATTTTATTTATTTGACAGATATATCAAAAAGTCATGCGTAGG 4208
Db 1050 rLeuGlyAspTyrArgAspIleLeuPhe----- 1059
QY 4209 AAGTAAACTRAAAGACAAATAGATTCCTTTCAAAAATGTGACCAAAAATCTCCTAA 4268
Db 1059 ----- 1059
QY 4269 TGGAAAAACACGCCAAGAAATGTGACAGAACATAGTATGATATGGAAGCTATGCT 4328
Db 1059 ----- 1059
QY 4329 AAGTGACAGTAAATAAATTTGGGCAAAAAAAGATGATTTTCCGAAAATCAGGTTACAA 4388
Db 1059 ----- 1059
QY 4389 CAACGTCAATTTAGTACAAAAAGCAACCACTTTGAGGAATTTGCCAAAGCACCAGTT 4448
Db 1059 ----- 1059
QY 4449 TTTACAGTGGCTAACCGAATGTGACGACGACATTTGCTATACAGCAAAAAATATTTGAA 4508
Db 1059 ----- 1059
QY 4509 GCGTGTGACGAGAAAAATGTAGTCAAAATGACCAATTTGAATGATACAGAAATTAATA 4568
Db 1059 ----- 1059
4569 GAATGCGAGACTACGTTAAATATATGAAAAAAGAGAGTGCATTCACAGATTA 4628
1059 ----- 1059
QY 4629 ATATTACAAGATAGACGACGACAAAAAAGATTCGATAGACACACATTCGTATATGCT 4688
Db 1059 ----- 1059
QY 4689 TACAGACTATCTGAAACGATCAACAGATTAATCTTGAACAGAAATTTACTCTACTTG 4748
Db 1060 -----SerGlySerAsnAspThr-----ThrservalSe 1069
QY 4749 TGGTATAGGCTGGAAGTCCCTGTGTGTACAAAGAAATATACAAATTTGTAGAAAAACA 4808
Db 1069 rLysAspThrProSerSerAsn----- 1077
QY 4809 GCGTTACTATAGTCCGACAAACATTTGTGGTGCACAAAAATTTATGAAAATGACGACAA 4868
Db 1077 ----- 1077
QY 4869 ATATTACTAATTTCCAGTAAAGATAGTCAAAAGATTAAGTAAAGAGCAACACACAG 4928
Db 1077 ----- 1077
QY 4929 TGGTATTAAGTGGCAAAACAAAGGTCTCTAATACTAATACTTGAAGAATTTGACTGA 4988

Db 1078 -----:||||| 1084
QY 4989 AGATGCTCTTTTCCTTCGTCGACACCTAATATGTTTCATGATCATGTGATGCAATTA 5048
Db 1084 ----- 1084
QY 5049 TACAGATCCAAAGTTAAAGATGAAATGGGTTGCCAAAAAGATTGATGAGAGTGGCGC 5108
Db 1085 -----LeuLeuAlaSerGlySe 1090
QY 5109 AAGGAAAGGTACAAATTTGGGTCAATCTACAAAAGAAAAAAGAAAGAAATTA 5168
Db 1090 rThrGlu-----GlnGluArgGluLysMet-- 1098
QY 5169 AACGTCGATGCCACACAAATATTCTTATGAGTCCCGCTTGATGCTATGAAATATAG 5228
Db 1099 -----AsnLysTyr----- 1101
QY 5229 TTTTATGATTTAAGAGATTAATTTAGTATGATTAATTTGAAAGATGAAAAACAAA 5288
Db 1101 ----- 1101
QY 5289 GACCGAGAAAAATTTGAGAAAAATTTTAAACAAAAATGCAACATCACTGGCAAGAGAG 5348
Db 1102 -----LysGluIleLysAsnPheArgLysCysSerThrGluArgSerAl 1116
QY 5349 TGATAGTACTACAGAAATCCCGTAGTACTGCCGAAATTTTTCGACAGCAAAATAA 5408
Db 1116 aProAsnLeuValSerHisPro-----GlnThrTyrTrpGluAsnAngl 1131
QY 5409 GGAATGTGTGTGACAGCAATGATATCGGGGTACAAACGTGTAGGAGATGGAATAG 5468
Db 1131 yLysTyrIleTrpHisGlyMetValCysAlaLeu----- 1142
QY 5469 TGGAAATAGTCAAGAAAGATGAAAGATCTAAAAAATGTGGTTCTGTACCTTCAGATGA 5528
Db 1142 ----- 1142
QY 5529 TGATTAATCTATGGGGAATAATCGCATGAAAGTACTGCTATCATGTTCTGTCAGTGT 5588
Db 1142 ----- 1142
QY 5589 TGCAGAAATGGGTGAGAGATTTTGCACAAATTAAGAAAGAAATTTGAGAAATTTGGTAGG 5648
Db 1142 ----- 1142
5649 GCGGTATATGATTAATCTGTGTGTATATGAAGATTAAGAAAGAAATGTACAGATGC 5708
Db 1142 ----- 1142
QY 5709 GTGTACAAATATTAATAAATTTATAGTAGTGAACACACAGATGAAAAACAAATCAA 5768
Db 1142 ----- 1142
QY 5769 AAAATATGTGAGAAATTAAGACAAATATATTTCCAGCATCTGTGCAAAAGATGACAGA 5828
Db 1143 -----ThrservalSplysile-----AlaLysGlyValGl 1153
QY 5829 GGAAGCTCGGAATATTTACAAACAAATTAATAAATTTGTGCAAAATTAATAAGTGA 5888
Db 1153 u----- 1153
QY 5889 TTGTGAATATAGTATATGAAGAAGATGTGCACACAGCATTAACGTATGTATATGCTCA 5948
Db 1153 ----- 1153
QY 5949 AAATATGCCGCATCATTAACAGATGAACCAAAAGAAAGTTGAAGAAAGTGAATTTGCTCA 6008
Db 1153 ----- 1153
QY 6009 AGTGCACAGAGTCCACACAGTGTACGAAGGAAACACAGTCAACAGGGATATCACTGAT 6068

```
Db 1153 ----- 1153
OY 6069 ATCAAAAGCGAGCGATCGAAAAAGCAAAAGAGCGCCCTACAAAAGCCGGA 6128
Db 1154 ----- 1154
OY 6129 AAAAGGAAATCTTAACACAGAAATGCGAGCACAAACGACCCGACAGCAGCACA 6188
Db 1157 nlysllelelunsn ----- 1157
OY 6189 ACMAACAGMAAACGACATCAACAGCAACAAACAGATCTGACGTGGGACAAATGCT 6248
Db 1161 ----- 1161
OY 6249 AAAGGCATCTCTTGAATTAACAGATAGCAGGGGTGGAATAGAGGTTGTAATCCAA 6308
Db 1162 ----- 1162
OY 6309 AACGATGACAAATATCTTAATGCGGTGTATGTAGTAAAGTCTAAAGAAATGAAA 6368
Db 1174 s ----- 1174
OY 6369 TGGCATATGATGCTCTCTAGAGAAAAAATATATATAATATACAAATATTTAA 6428
Db 1175 ----- 1175
OY 6429 TTATGAACGTGAATAAAGCGTGCATGATATATAAAGCGCTTTATTAATGTGAGC 6488
Db 1182 nvallyleulasp ----- 1182
OY 6489 AATGAAATCTCAATTTTGTGCTTAATAATATATTAATGAAATCTGACAGCAAAATGA 6548
Db 1187 ----- 1187
OY 6549 ATTGCAAAATGGAACATCTCCAGATGATTAATAAGAAATATATATATACATATGCTGA 6608
Db 1189 ----- 1189
OY 6609 TTATGAATATGTTTTTGTGATGATATCTAATGATTAATAATATATATATATATAT 6668
Db 1198 ----- 1198
OY 6669 AAATAGTGTACAAACCTTCTCAATGATTAATAATAAATAAAGAAATTAATACTGTAAC 6728
Db 1204 ----- 1204
OY 6729 TGAAGATTTACGTAATAATATTTTGGAGAAAAATTAATTTATTTGGAGAAATGAT 6788
Db 1208 ----- 1208
OY 6789 AATGGAATTAATCTATCATCTCAGACAGMAAACGMAAAAGMAAAATTAAGATTAATTA 6848
Db 1208 ----- 1208
OY 6849 CCAGTACATGACATGACCAAACTGACGCCCTTCCCTGAAGATTTGTAATAAAGCCCA 6908
Db 1209 ----- 1209
OY 6909 AATTTGAGATGTTCAACAGATGGGAGAGAAATTTGTATATAGAGAAAGAACATCTT 6968
Db 1217 rlyrpheargtprpheluglultrpelyglulserphelcyaragluarqlyslsaryle 1237
OY 6969 GTTAAATTTGGAGCGGCTGTAGGAATATGAGTGTATGCT 7013
Db 1237 ulysglinlelyslaspcyslyslvalglu ----- 1237
OY 7014 TATAGCGTGAACAACAAGAA ----- 7014
Db 1235 rgllyspelyglulalacysaspserilserthhisasptyserthrvalproserph 1275
OY 7036 ----- 7036
Db 1275 eamncysprogllycysglyslshiscyaserSertyrarglystrpiledgluarglysl 1295

OY 7083 AACTGAATATGAAAGACAAAGAGAAAAAGTTCAAAAAGATTAAGATGCAAAAAGTATA 7142
Db 1295 silegluphehislyslserasnlatyrglyglu ----- 1307
OY 7143 GGATTAATCTCTTAACGTAAGAGACATAGAGAGGCAACATGCTCATGATATTTAA 7202
Db 1308 ----- 1308
OY 7203 CATGAATTAAGAAATTAATGTCGCAATTAAGATGTCTCTGTATGCAAAAACCTTCTTC 7262
Db 1318 ntrhpheasplysluphecy ----- 1325
OY 7263 ACAACTACCAAAACAAACACACACATCAATCATCATCCGATGCTAATGATATGCCAATTC 7322
Db 1326 ----- 1326
OY 7323 GCTGATTAATGTTCTCTGAAGAAATTTAACAAAGTGTGAGTCTCTGAACCTTCAAAAAGG 7382
Db 1327 rleuglultrphtprasp ----- 1333
OY 7383 ATCTATGATTCATACAAAAAATTAATCTGAACTTAATACCTATGAATTTGTAGAGA 7442
Db 1333 ----- 1333
OY 7443 ACAGCATATTAATTAATTAAGAAAGAGAAATATATATGATATACCTTGAAGGAAA 7502
Db 1334 -AlaAlalyspHeleularg ----- 1340
OY 7503 AATTAATCTTAATGAGTCTACAAAGAAAGAAAGAAAGTAAATAGTTGACTAATTAATA 7562
Db 1341 ----- 1341
OY 7563 TCCCTGCGATCTCTAAGAAACCTTATGACCTGATTAATATATAGGAAGAAACCTTG 7622
Db 1344 yprocylystrhsanlyslutryglylyasp ----- 1355
OY 7623 TGAATAAGAGAAAGAAATCTGTTTAAGTATGATTAATGAAATGTTACAAATATTC 7682
Db 1356 ----- 1356
OY 7683 AAAGTCTATCAGAGAGAAAGAAAGATATGTCATCT ----- 7739
Db 1363 rlystrhpheglinhstrhglu ----- 1382
OY 7740 CTTAAGAAATTTAGATGAATTAATAATTAAGAACTTAAGATATATATCTCTAAA 7799
Db 1382 s ----- 1382
OY 7800 AATGTTCTGCAACTGCACGAAATGAAGAAATAGACATTAATAAATCAACTCAGA 7859
Db 1383 ----- 1383
OY 7860 GAACGGG ----- 7890
Db 1383 nasnclyasnrclyslvalserglyleuasnghlysnrclyslaspspserilseas 1403
OY 7891 ----- 1403
Db 1403 palalyglulalalysmelarigserSerthrtraspval ----- 1421
OY 7935 AACGATATGTTACGAATTTGCTGTTACTTCCCTCGTAGAAATTAATAATTAATAGCT 7994
Db 1421 lserasp ----- 1423
OY 7995 TTTTGAATACATATATGAAATATGAGAAATATAAATAAAGTAGAAATTAATACACGA 8054
Db 1424 ----- 1425
OY 8055 TGTAACAAGCTT ----- 8091
Db 1425 prhrasntrhphegluglyaspaspheulaysaspalacysglinhslalasnlllehele 1445
```

OY	8092	-----AGAAAGATTTTGTGAACAAATGACGTGCACAAAGACACGAAGAATGCACAAACT	8144
Db	1445	sglyllletrgyasppalltrpyscys-----	1454
OY	8145	TTTTAGAAAAGGAAGATGGATGATTGAA-----CGCATATCATTAATACAGATNA	8198
Db	1455	-----glylyrvalcysgiyalaspliecysgluinthrashlleasnguuar	1471
OY	8199	GTGTGGACANTAGACGACGATCCACCTGTTGATGATTATTAATCTCAACGGTTT-----	8250
Db	1471	gThraspglylys-----gluyrileglinllearghlaLeuphely	1485
OY	8251	-CGATGGATGACATGATAGTCGTCAATATTAATTTGTAAGACACTGATGGAAGAATTGGA	8309
Db	1485	sargtrpyvalgluasnhe-----Leugluaspyrasyly	1497
OY	8310	ATTTAAAAAATCATGTCATCTGT-----AAACATCTGCACATGACGACGAAGATNA	8360
Db	1497	sileasnasplyslleserhsicyslllelysglysgluiglyserlysCysllleasncl	1517
OY	8361	TTATATGAAAATAGTAGTCGACACGTGTAAAACGAGATGTCAAGATATAAAAATTTTGT	8420
Db	1517	ycysglulysasn-----serlyscysleugluistrpliegulyslys-----	1533
OY	8421	TCTTAATTAAGAAATCTCTATTCATATACATCAATTAATTAATCAAAAGATTCATGAA-	8478
Db	1533	ealglutrrpeluasnle-----lyslasasgphasnaspelntlyrgluas	1549
OY	8479	-----CAACCATATATATCAAAAATCTCTACTTATGATCATGTGTCAAAATTTGTGACA	8531
Db	1549	nlyasapelnrproaspyrtrsnvallyserlle-----leugluinleullelpr	1566
OY	8532	AAAGTTGAAAACCTTTTAAAAGTAAGATGTCCTGTGAGACCTTTCTGAATATCTTCATGA	8591
Db	1566	olysillealavalalnasapelnaspsanvallelyleucysvalpne-----gl	1584
OY	8592	AACAGTNAAGCTTTGCAATATTAATTAATTAATGAATAATGATGGTCTTCCAAATATACGAC	8651
Db	1594	uasnsertlysclycysThrleulleaseranthrgluasnasnlygluasn-----	1601
OY	8652	ATATGCTTTGAGAGAAACACCAAAAAGTTATAAGAGAGCTTCAGCTGACACTCTTC	8711
Db	1602	-----aspalaillleaspcysmelleullyly	1610
OY	8712	TAGAAAT--CCATTTGATTAATTTGT-----CCTACCGATTCACAAACMAAGATGATG	8759
Db	1610	sleuglyvallysalhalysasnycspProgllyspProsergllygluylglnserhsapcy	1630
OY	8760	TAGAGATTAACAACCTTTTACCTTCCTGCTCGAAGATGATTTTGAATATATCTTGATNA	8819
Db	1630	slysgluprproProleu-----Proaspgluinlaspelnasn-----	1643
OY	8820	TTGAGACGATACCTGTCTTCTTAATAGTTCAGATGATTAACAAGAGTATTAATTCCTCC	8879
Db	1644	-----ProgluiguasnthrlleugluinPr	1652
OY	8880	AAGAAGAGACATTATATGTACAAAGACCTATCACTGCATTAATTAATAGAAAAGTATNA	8939
Db	1652	olys-----PheCysPro--ProthrthrlnPrProgluigluLysGly-----	1666
OY	8940	AGAAATTTTAAAAAAAACCTCTTACTCTGCTTCAGTCAAGGACAAATTTTGAGTCA	8999
Db	1667	-----glyluluthrCysgllyas	1672
OY	9000	AAATATTAATCGGAGAAAGTGTGCTTTGAGGCAATGAATATAGTTATGCAGATTA	9055
Db	1672	nlysgluinlulysysasplu-----	1679
OY	9060	TTCCGATATATTAAGAGACGATATGATGACACTTCATTATCTGAAAAAATTAANA	9119
Db	1680	-----Lysly	1681
OY	9120	AATATTTGAACATCAATGACAGAACCGAAATCTGAATAACATGCTGGGAAATTAATNG	9179

[illegible]

QY	97	AAAGGACAGAAAGTTTGGAAAGCTTATGCCAAAATATA	15	---AGACATCATCAAAA
Db	13	GlusterVallysAspIeuphneAspHqIleGlyLysValTgylulysThrGluLys	32	
QY	154	PTATGGAAGAAAGCAATGTCATGCTGTTGAAAGGCGATTTGACGAACGAATTTTCGGGT	213	
Db	33	IleAlaLysaLgTyrThrThrGluLeuHisSgLYAspLeuSerLysAlaThrTyr	50	
QY	214	GGTCTCTTACGCCACGTAATTAAGCATTAATTATTATTATCA	255	

Dd	51	-----				-----	ProGluCylserThrGluAsn	62
Qy	256	TATCCAGTAAATTAGTACATAAGCAACTACTAAATTTCAGGATGATGATGTAATTGG				-----	ProGluCylserThrGluAsn	62
Dd	63	AsnProCysLysLeuGlnIleuTyrAspTyrAsnThrAsnValThrHis				-----	GlyheGlyGln	81
Qy	316	AGACATCCCTGGCCATGGTGGAGAACAAACCCGTTGGATGAGATGAGAAATCTGAATGT				-----	GlyheGlyGln	81
Dd	82	GluTyrProCysGluThrAspIleValGluThrPheSerAspThrGluGlyAlaGlnCys				-----	GlyheGlyGln	81
Qy	376	GGAATTAATAATACGTAATTATAAGAAAAAATGATGCTATAGCTGTGGCCACTAGA				-----	GlyheGlyGln	81
Dd	102	AspLysLys-----LysIleLysAsnSerGlnGlyAlaCysAlaProThrArg				-----	GlyheGlyGln	81
Qy	436	AGACGACATATGCTGATATAAAACCTGGAGAGCTCAATAGATATAAT--ACCCAAAT				-----	GlyheGlyGln	81
Dd	119	ArgLeuIleValCysValArgAsnLeuGluAsnIleAsnAspTyrSerLysIleAsnAsn				-----	GlyheGlyGln	81
Qy	493	ATTCAATGATTTATTTGGAAATGTAAGTACTAGTACAGCAAAATACGAAGTGATCAATTGT				-----	GlyheGlyGln	81
Dd	139	LysHisAsnLeuIleValGluValCysLeuAlaIleLysTyrGlnGluGlnSerIleThr				-----	GlyheGlyGln	81
Qy	553	AATAATCATACCA--CATTAAGAACT-----TCAGACGCTTACTGCT				-----	GlyheGlyGln	81
Dd	159	GlyArgTyrProGlnHisGlnGluThrAsnProAspThrLysSerGlnLeuCysThrVal				-----	GlyheGlyGln	81
Qy	595	CTTGCACGAAGTTTTCAGATATAGTATATGATATATTAAGAGAAATAGATGTTTAA--				-----	GlyheGlyGln	81
Dd	179	LeuAlaArgSerPheAlaAspIleGlyLysIleLeuGlyLysAspLeuTyrArgGly				-----	GlyheGlyGln	81
Qy	652	-----CCAAATGCTCATGACAAAGTAGAAAGGGGTCCGAGAGTTTC				-----	GlyheGlyGln	81
Dd	199	GlyAsnThrLysGluLysLysLysAspTyrGlyLysLeuGlnGluAsnLeuLysThrIlePhe				-----	GlyheGlyGln	81
Qy	697	AAGAAATACATGATGATGATG-----GAAATGAAAGTAAATAAT				-----	GlyheGlyGln	81
Dd	219	GlyHisIleTyrAspGluLeuLysAsnGlyLysThrAsnGlyGlnGluGlnGlnLys				-----	GlyheGlyGln	81
Qy	736	GATTCAATCCCGATGATCTCGAAATTTATTAATTAAGAAAGCATGCGTAATGTG				-----	GlyheGlyGln	81
Dd	239	ArgTyrArgGlyAspLysAspAsnAspPheTyrGlnLeuArgGluAspTyrPheAspAla				-----	GlyheGlyGln	81
Qy	796	AATAGAAATTAAGATGGGAAGCATATAACATGTAATGCA--TCATTAATATGGAATAT				-----	GlyheGlyGln	81
Dd	259	AsnArgGluThrValTyrPheLysIleIleThrCysAsnIleAspLysTyrGlnTyrSerGln				-----	GlyheGlyGln	81
Qy	853	TTTATGCAATCAGAAAGTAAATACACCATTAATTTTCAATCTTAATCGGCGCCATAAACAA				-----	GlyheGlyGln	81
Dd	279	ProThrCysGlyArgGlyGlnIleProTyrValThrLeuSerLysCysGlnCysIleAla				-----	GlyheGlyGln	81
Qy	913	GGAAGGTTCTTACCAATTAAGATTAATGTCGCCCAATATTTACGTTGGTTCGACGAATGG				-----	GlyheGlyGln	81
Dd	299	GlyGluValProThrTyrPheAspTyrValProGlnTyrLeuThrGlyTyrPheGlnGluTyr				-----	GlyheGlyGln	81
Qy	973	GGAAGAAGTTTTCGCCAAAAAGAAATATTAATTTGAAAAAGGTCAAGGACTCCGTGCT				-----	GlyheGlyGln	81
Dd	319	AlaGluAspPheCysArgGlyLysLysLysIleProAsnValLysThrAsnCysArg				-----	GlyheGlyGln	81
Qy	1033	-----AATGACAAAGAACCGTTATATGTAATGATGATGACATGTAATGTAAGCA				-----	GlyheGlyGln	81
Dd	339	GlnValGlnArgGlyLysGlnLys--TyrCysAspArgAspLysLysAsnCysAspGly				-----	GlyheGlyGln	81
Qy	1084	ACTATTTGAAAAAAGGTATTTTTCGATTTGGATTAATAAGTGTATGATGCTGTTGCATTA				-----	GlyheGlyGln	81
Dd	358	ThrIleArgLysGlnTyrIleTyrArgLeuAspThrAspCysThrLysCysSerLeuAla				-----	GlyheGlyGln	81
Qy	1144	TGCAAAGTTTGAAGTTGGTTAGGAATCAACAAGAGATTTTAAAAACAACAAAGAA				-----	GlyheGlyGln	81
Dd	378	CysLysThrPheAlaGluTyrPheAspAsnGlnLysGluGlnPheAspLysGlnIleGln				-----	GlyheGlyGln	81
Qy	1204	AATATGAAAAAATATACATATTTTATGCAACGATACAAATTTGTCAATATATAT				-----	GlyheGlyGln	81

Db 398 LysTyrGlnAsnGlnIleSerGlyGlyGlyArgArgGlnLysArgSerThrHisSer 417
QY 1264 AATAGTGA-----TATTATAACAATTTTATGAAAACTTAAAGAAAGCAATAT 1314
Db 418 ThrIleGlyGlnGlyGlyGlyGlyHisPheAsnGlnGlnIleuLeuArg----- 433
QY 1315 GCACATAATGAC-----ACCTTTTAAATTTACTAAATGACGAAGATAT 1359
Db 434 -----AsnGlnIleLysAspValArgSerPheLeuGlnLeuSerLysGlnLysIle 451
QY 1360 TGTAAAGAGAGATACCA---GGAGAAAAGATATTACTTACTTACACAGTGTGATGAC 1416
Db 452 CysLysGlnArgIleGlnIleValGlyGlnGlnIleThrAlaAsnTyrGlnAsnPheGlnAsnGln 471
QY 1417 AAGGATATTATTATGCTTCAGATATTTGCCAAGTGTGCTCCAGTGGGGGTCAATGT 1476
Db 472 SerAsnThrPheSerHisThrGlnTyrCysAspArgCysProLeuGlyGlyValAspCys 491
QY 1477 GATGATATAAATACACACCAATCAGATATGAT---CGTGAAGCTGTAAATATGAA 1533
QY 492 SerSerAspAsnCysArgLysLysProAspLysSerCysAspGlnGlnIleThrAspLys 511
Db 1534 GACTATAAACCTTCATGGGGGTGGAAGCTACTAATATCAGTCTTATAGTGTAAT 1593
QY 512 GlnTyrProProGlu-----AsnThrLysIleProLysLeuThrAlaGlnLys 528
QY 1594 GAACAAGTGTATTTACACAAAAATAGAAAATTTTGTAAAGCTCAACATTAATACAA 1653
Db 529 ArgLysThrGlyLeuLeuLysLysTyrGlnLysPheCysLysAsnSer----- 544
QY 1654 GATTAATAATAT-----CAAAAATGGGATGCTATAT-----AAG 1689
Db 545 AspGlyAsnAsnGlnLysGlnIleLysLysTyrGlnLysHisTyrGlnLysAsnAspLys 564
QY 1690 GATGAA-----AATATAATAGATGT---AACTGGAACAAAATATGGAATCAAT 1737
Db 565 AspAspGlnLysAsnLysPheAsnAsnCysIleGlnGlnAspTyrThrSerLysAsn 584
QY 1738 AATGATATATCTTAAGATATATCATTTTCAATATTTTGAATTAAGGTTACATATTA 1797
Db 585 ValTyrTyrPro-----IleSerTyrTyrSerPhePheTyrGlySerIleAspMet 602
QY 1798 TTAAGGATATCTTAAAGTGAATGACAAACTTAAATCTGTATTAATATACAAACAG 1857
Db 603 LeuAsnGlnSerIleGlnTyrParGlnArgLeuLysSerCysIleAsnAspAlaLysLeu 622
QY 1858 ---CATTTATGATGAATGTATACAGAAATGCTTATGTTTGAAGATGGTTAAACAA 1914
Db 623 GlyLysCysArgLysGlyCysLysAsnProCysGlnCysTyrLysArgTyrPheGlnLys 642
QY 1915 AAGAAAGAAAGATGGAATAGTATTAAGAAACCTGTCACAAAAAAAGAAATATACAGCA 1974
Db 643 LysLysAspLysPheLysPheLysIleLysGlnPheArgLysGlnLysAspLeu----- 660
QY 1975 TCGTATTATAGTAAATATTAATCTTTTGAAGCTTATTTTAAAGTTATGATTA 2034
Db 661 -----LeuLysAspIleAlaGlyMetAspAlaGly----- 670
QY 2035 CTTCAGAAAGATGAACCAAAATGAAAACCTTATGGAATATTAAGAAAAAATAAT 2094
Db 671 -----GlnLeuLeuGln----- 674
QY 2095 GACTTTTCCAAATTTGAAAAATTAAGGACTATTTAGAGATGCAATAGAACCTTGTGTA 2154
Db 675 -----PheTyrLeuGlnLys-----IlePheLeu 682
QY 2155 GATCATTAAAGAACTGCCAGATATGTAAAGACATATATCAAAAGATGTGAA 2214
Db 683 GlnAspMetLysAsn----- 687
QY 2215 ACATCCATATATGCAACAAACCGGTGTGTTAAACCTGTGAGGACGCAACCCACT 2274
Db 687 ----- 687

QY 2275 AAAAAATATAAAGAAATAGCACATACTTTAAAAAGAGTGCATACGAGAAAGCAAGAAAT 2334
Db 687 ----- 687
QY 2335 CGTGTCTTCATTAATTTGAAGAGAAAGGACACAGAAAGGTATATATAACGTGGGGTAGG 2394
Db 687 ----- 687
QY 2395 AGAAAGACTTCAGACAAATTTATGTAGAAATATGATAAATCTTATCTATCTT 2454
Db 687 ----- 687
QY 2455 GGTTTTCAATGAGACCAGTGTATGCCAAGGACAGGTGATATACAAAGATTT 2514
Db 688 -----AlaAsnGly----- 690
QY 2515 GTGCTGAAGACTGAATGGGAAGTGAATCCGGAACACATGCGTAAAGATCAAGAAAGTGT 2574
Db 691 -----AspPro----- 692
QY 2575 ATTATGCTCTCTAGAACACATATATGTATCATCCAAATTTGGAACATTTACAAAGCAT 2634
Db 692 ----- 692
QY 2635 GATACCCACTTAATGTAATATTGTTGATGATTAATTAATCTTTTGGGGGAT 2694
Db 692 ----- 692
QY 2695 GTTCTTCTATCAGCAAAATATGAAGCAAAACAGATATACGATGTATAAGAAAGAT 2754
Db 693 -----LysValIleGlnLysPheLysGln----- 700
QY 2755 AACCTAAAGGGCCCAAGAAGTAAGTACACCCAAACACAGACAACTATCTGTGAGCT 2814
Db 700 ----- 700
QY 2815 ATACGTACAGTTTTCAGATATATAGTATATTAATTCGAGAAAGATCTGTGGAAAGA 2874
Db 700 ----- 700
QY 2875 AACGGTGCATGTGTAAGCTGCAAGACATTTGGAACCTGTTTGTGTAATATACATAG 2934
Db 700 ----- 700
QY 2935 TCACCTCAAGCAAGCAAGAAATGATTAATATATGATGATGCCCCCAATATTTAAATTG 2994
Db 701 ---IleLeuGlnLysGlnAsnGlnIleValGlnAspPro----- 712
QY 2995 AGGAAAAATTTGTGGAGAGCTAATAGACCAAAAGTATGGAAAGCATGAATGTGATATA 3054
Db 712 ----- 712
QY 3055 AATATTTGAAGATTAATGCGGACACACCAATCAACAAAGTACTTATTCGGATATAGT 3114
Db 713 -----LeuLysThrLys----- 717
QY 3115 GATCATACACCATGTGATGATATATCCACAAAATTAAGATGATGATGACCAATGGCA 3174
Db 718 -----ThrIleAspAspPheLeuGlnLysGlnLeu----- 727
QY 3175 GAATGTACTGCAAGGTGCAAGAAAAAGAGTATGATTAAGTTGAAGAGAAAGTGAAGAG 3234
Db 728 -----AsnGlnAlaLysAsn----- 732
QY 3235 TGTAAAGATTAAGATATATGCTCAAGGCTGTACGAAAGAGAGTGTACAGTGTGACAG 3294
Db 733 CysValGlnLysAsnProAspAsnGlnCysProLysGlnLysAlaProGly----- 749
QY 3295 TGCACAGAACTGTATATGATATATATGATATATATGATTAAGAAAGAAATGGAAT 3354
Db 749 ----- 749


```
QY 3355 ATAAATTCAGATTAATACAAAGAAATACATGACACACAAATGCTGTAGTAATAGT 3414
Db 750 -----Asp 750
QY 3415 GGTATGTAAGCTCCAGTACTGCCAAAAATCATATAGACAGAAATGTTATGAAATTTTG 3474
Db 751 GlysAlaIaIaProSerAspProProArgGluAspIleThrHis -----764
QY 3475 TCGGAATTAATACCAACAAATGTCGCAAAAAGTAATAAGTGTACTAGTGAAGT 3534
Db 764 -----764
QY 3535 GCCTGCTACTGTACTAACACACAGTATGAAAATGTTGAGCATATCTCATGATACAGA 3594
Db 765 -----HisAsp-----766
QY 3595 AATTGTGATGATGTCAGTACACAAATGAGTTTGTGATGAAAAGATGATGTAAGAT 3654
Db 767 -----GlyGluHisSerSerAspGluAsp 774
QY 3655 AAGCAAAATATATGCTTTAGATTAACACAGGACCATGATGCTGCTGTGTA 3714
Db 775 GluGluGlu-----GluGluGluGluGluGluGluProProAlaGluGluThrGlu 791
QY 3715 AGTGA-----TCGAACCGACACAGGTAACAGATTAACAAAAAGGCAAGAAAG 3771
Db 792 GluGluGluGluGluSerGluSerGluValValGluGluGluGluProGluGlu 811
QY 3772 GATACGAA-----TGTAACACAGTAAT 3795
Db 812 AspThrGluIuThrValProThrThrProThrValAspValCysAspThrValIys 831
QY 3796 GATATCTTAAGAAACAGATGAAAGAAAGAAAGTAAGTTCATCCAAAAAGAT 3855
Db 832 ThrAlaLeuAlaAspThrGlySerLeuAsnAlaAla-----CysSerLeuIuThrVal 849
QY 3856 AGTAATGATATCCGATTCGCAATGC-----GGAATATTAATTAAGTGAAGAC 3906
Db 850 ThrGlyLysAsnThrGlyThrArgCysIleAlaProSerGlyThrThrSerGlyAsp 869
QY 3907 CCTCGTGTGTATGCCCCCTAGACAAAGTAATGCGTACATTTCTGGCAATGAT 3966
Db 870 GlyAlaIleCysValProProArgThrGluGluLeuGluCysLeuIuThrLeu-----886
QY 3967 AATGAATTAATAAAATTTACA-----TCACAAGTTAATTAAGAACTTCATCAA 4020
Db 887 -----LysGluLeuSerAspThrThrGluLysGlyLeuArgGluAlaPheIleIys 903
QY 4021 TCTGCACGACGAAACATCTTCATGATTTATTATTA-----AGTAAGATGATGA 4077
Db 904 ThrAlaAlaGluGluThrThrLeuLeuThrGluLysGlyLeuAspLysGluAsnGlu 923
QY 4078 -----GGAATGAACTGAT-----AAGAATTAAGAAAGCGCAAAAT 4116
Db 924 ThrAlaSerThrGluLeuAsnIleAspAspProGluThrGluLeuAsnGlyGlyGluIle 943
QY 4117 CCTCCGCAATTTTGAATTCATGTTCTACACATTTGAGATTTATGATTTTATT 4176
Db 944 ProGluAspPheLysArgGluMetThrThrPheGlyAspThrArgAspPheLeu 963
QY 4177 GGAACAGATATATCAAAAGTCAAGTGAAGAACTAATAAGACCAAAATGATCT 4236
Db 964 GlyArgThrIleGly-----AsnAspLeuAspLysValAsnAsnAlaIleThrAla 980
QY 4237 CTTTCAAAATGGTGACCAAAATCTCTAATGA-----AAAAACGCCAA 4284
Db 981 ValPheGluAsnGlyAspHis-----IleProAsnGlyGluLysThrAspArgGluArgGlu 999
QY 4285 GAATGCTGACAGAAACATAGTATGAGATATGGAAGCTATGATGCTAGTATAA 4344
Db 1000 GluPheThrGlyThrThrGlyLysAspIleThrLysGlyMetLeuGlyAlaLeuGluGlu 1019
QY 4345 ATGGGGCAAAAAAGATGATTTTACGAAATACAGGTTACAAACAGCTCAATTTAGT 4404
Db 1020 AlaGlyGlyLys-----ThrLeuThrGluThrThrAsnIuSerAsnValIlePheAsn 1038
QY 4405 GAC-----AAAGACACATTTGAGGAATTTGCCAAACGACCCAGTTTATGATGG 4458
Db 1039 GlyHisLeuThrGlyThrLysLeuAsnGluPheAlaSerArgProSerPheLeuArg 1058
QY 4459 CTAAAGCAATGTCACACACATTTCTATACACGCAAAATTTTGAAGATGTCAG 4518
Db 1059 MetThrGluThrPheLysAspIlePheCysArgGluArgIleThrGluLeuGluIleLeuLys 1078
QY 4519 GAAAAATGTAAGTCAATATGACCAATTTGAAGTGTGATACAGATGTAATAAGAAATGCCAG 4578
Db 1079 GluArgCys-----1081
QY 4579 GACTACGTTAATATATGAAAAAAGAGTGATTCACAGATATATATTTACAAG 4638
Db 1081 -----1081
QY 4639 GATGAACGCAAAAAAAGATTCGATAGACACACATTTGTTATGTTACAGACTAT 4698
Db 1082 -----MetValIuGluThr 1086
QY 4699 ACTGGAACGATCAACACATTTACTTGAACAGAAATTTACTGTAAGTTGTGTGATAG 4758
Db 1087 AsnGly-----1088
QY 4759 CCTGGAAGTCTGTGTGTGTACAAAGAAATATACATTTAGAAAAACAGGCTTACTAT 4818
Db 1088 -----1088
QY 4819 GATGCCGCAAAACATTTGTGGTGACAAATTTATGAAATGACGACAAATATACATAAC 4878
Db 1088 -----1088
QY 4879 ATTGGAATTAAGATTAAGTGCAAAGATTAAGTAAGAGGCAACACAGGCTATTAG 4938
Db 1088 -----1088
QY 4939 TGGCAAAACAAAGTCTAATTAATACAAATTAACCTTGAAGAATTTGACTGAAGATGTCTT 4998
Db 1088 -----1088
QY 4999 TTTCTGTGTGTGACATGATATATGTTTCATGATGATGATGCAATTTATACAGATCA 5058
Db 1088 -----1088
QY 5059 GAATTAAGATGAAGATGGTTGCGAAAAAGATTGATGAAAGTGGCGCAACGAAAGG 5118
Db 1088 -----1088
QY 5119 TACAATTTGGTCAATACTACAAAGAAAAAAGAAAAAGAAAAATTAAGCTGGAT 5178
Db 1088 -----1088
QY 5179 GCGCAAAATTTCTTATGAGTCCCGCTTGTAGTCTATGAATATATGATTTTATGAT 5238
Db 1088 -----1088
QY 5239 TTAAGATTAATTTAGTATTTGATTAATTTGGAAGATGAAAAACAAAGACCGAGAA 5298
Db 1088 -----1088
QY 5299 AATTGAAGAAATATTTAAACAAAAATGAAACATCACTGTGCAAGAGAGATGATAGT 5358
Db 1089 -----AspLysGlyLysAsp-----1093
QY 5359 ACAGAAATCCCGTACTACTGCGCAAAATTTTCTGGAACGAAAAATAGAATGTGTG 5418
Db 1093 -----1093
QY 5419 TGAACGCAATGATATGCGGGTCAAAACGTGTTAGGATGATGGAATATGTAATAGT 5478
```


QY 7636 GAAATCGTTTAAAGTAGATTATGATGAAATGTTACAAAAATTCAAAGTCTATCAG 7695
Db 1220 ----- 1220
QY 7696 GAGAAAAAAGATATGTGTACTCCAAAGAGAGACATATGCTTATAGAAATTTAGAT 7755
Db 1220 ----- 1220
QY 7756 GAAATTTAAATGAAAGACTTAAAGATAGTAATATCTCTTAAAAATGTTCTCGAAGT 7815
Db 1220 ----- 1220
QY 7816 GCAGAAATGAGAAATAGACATATATAAAAACTTCAACTCAGAGAACGGGTCCGCAATG 7875
Db 1220 ----- 1220
QY 7876 AATCCAAATATGTACTATAGAAATATAGTTTCTGCTATCTGGGTGACATAGTTAGAGA 7935
Db 1220 ----- 1220
QY 7936 ACAGATATGTACGAATTTGGTGGTTACTTACTTCCCTCCGTAGAAATTAATATTAAGTT 7995
Db 1220 ----- 1220
QY 7996 TTTGAATACATATATGAGAAATGAGAAATATAAAATTAAGTAGAAATATACACAGAT 8055
Db 1220 ----- 1220
QY 8056 GTTCAAAAGTTTCTGCTGCTGTGGGAGCTAATAGAAAGATATTTGGAAGCAATG 8115
Db 1220 ----- 1220
QY 8116 ACGTGCAAGCACCAGAGAGATGCAAAACTTTTNGAAAAAGAAATGATGATTTGAA 8175
Db 1220 ----- 1220
QY 8176 CGCATACATTAATACAGATAGATGTGACATAGAGAGATCCACTGTTGATATAT 8235
Db 1220 ----- 1220
QY 8236 ATACCTCAAGGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 8295
Db 1220 ----- 1220
QY 8296 GAGAAATGGAATAATTTAAAAATCATGATGATGATGATGATGATGATGATGATGATG 8355
Db 1220 ----- 1220
QY 8356 AATGATTATGAGAAATATAGTGTGACAGCTGTAAGAGATGTCAGAAATATATAAAT 8415
Db 1220 ----- 1220
QY 8416 TTTGTTCTTAATGAAATCTCTATTCGATATACAAATCAATTAATTAACAAGATTTGAT 8475
Db 1220 ----- 1220
QY 8476 GAAACAACATATATACAAAAATCTTACTATGATCATGTTCAAAATTTTGTACAAAG 8535
Db 1220 ----- 1220
QY 8536 TTGAAACTTTTAAAGATGATGTTCTGTGAGAGCTTTTCTGAATATCTCATGAAACA 8595
Db 1220 ----- 1220
QY 8596 AGTAGTGTGAATATATAATTTAATGAATAATGATGTTCTTCCATATATACAAATAT 8655
Db 1220 ----- 1220
QY 8656 GCTTTCGAGAAACCAAAAAAGTTATTAAGAGAGCTTCAGTTGTACACTACCTCTAAG 8715
Db 1221 ----- 1223
LeuProlys ---
QY 8716 AATCATTTGATTAATTTGCTTACGATCAAAAAACAAGATGATGATGATGATGATGATG 8775

Db 1223 ----- 1223
QY 8776 TTTACCTTCTGTCGAGAAATGATTAATTAATTAATCTGATTAATTTGAAACGATACCTT 8835
Db 1223 ----- 1223
QY 8836 GTTCTTAATAGTCAATGATGATTAACAAGGTGATTTGATTTCTTCCCAAGAAAGACATTAA 8895
Db 1224 ----- 1226
ProProlys ---
QY 8896 TGTACAAGACCTATACATCATTAATTAATAGAAAAAGGTATTAAGAAATTTAAAAAAA 8955
Db 1226 ----- 1226
QY 8956 AAACCTTCTTACTTCTGCTTTCAGTCAAGACAAATTTGTTAGTCAAAAAATTAATTAATCGGA 9015
Db 1226 ----- 1226
QY 9016 GAAGAGTGTGCTTTGAGGCAATGAAATATAGTTATGAGATTAATTCGATTAATTAATA 9075
Db 1226 ----- 1226
QY 9076 GGAAGTATATGATGACACTTCATTAATCTGAAAAATTAATAATTAATTTGAACATCA 9135
Db 1226 ----- 1226
QY 9136 AATGAAGCAACGGAATATCTTAAAAATGATGAGGAAAAATTAAGAGTCAATATATGAC 9195
Db 1226 ----- 1226
QY 9196 GCTATGTAATGTGATTAATAATTTGCTACTTCAAAAGTAACATTAATGATGAAGATGCTGT 9255
Db 1226 ----- 1226
QY 9256 CAATTCAGAAAGATGAAAGAACTAATCAATGTTCTTCTGTTAATTAATGAGGCAAG 9315
Db 1227 ----- 1229
LeuProlys ---
QY 9316 CAAGCATGTAAGAAAGAAACATGTAATGATTCATTAATAAACAAATGCTCTGTTCA 9375
Db 1229 ----- 1229
QY 9376 AACGAGTAATTTTGAAGCTCAGAAATTAATTAAGACAACTGGATGACAAATGATAT 9435
Db 1230 ----- 1232
ArgInPro ---
QY 9436 AGAAAAATATTAAGCTTGAATATATGATTAATAAATACAAATGAAAAATCTAATATATAA 9495
Db 1232 ----- 1232
QY 9496 TATAAGCAATTAAGAATCAATCTTCAGTATTAATTAAGCAATTAAGCAATGGAAGAAAT 9555
Db 1232 ----- 1232
QY 9556 GTTACGATATTAATAAATCAAAAGATTCCAATGCCCTTGGAGTTAATGATTAAT 9615
Db 1233 ----- 1236
LysGluArgAsp ---
QY 9616 GAAATAGTTACAGAGAACAAAAATTAATGAATAATTAATCAAGAAAGTACTAAAAAAA 9675
Db 1236 ----- 1236
QY 9676 TTATATCTGTTATTAATTTGTTGAAGATGAACACAAAAATCATGTAATAGATGA 9735
Db 1236 ----- 1236
QY 9736 AATTAATAAGAGAGCAACAGTTCTGCTTAAGCACTGATTTCTTACACCCCAT 9795
Db 1237 ----- 1241
PheProThrProAla ---
QY 9796 GTAGATCTTCTATCAAGACCTTTATCTCAACATGAGATGATGATGATGATGATGATG 9855
:::

Db 1242 Leu-----1242
 Qy 9856 AAAATGATATATGAAGTAGATCTGTTATTTGATCGCGTATGATA 9915
 Db 1243 LysAsnAlaMetLeuSerThrIleMetTrpSerIleGlyPheAlaThrPhe 1262
 Qy 9916 GCGCTTCATTTCTGAGAAAAAATTCMAATGCTGTGGACTGTGGCTATGCAAT 9975
 Db 1263 ThrTyPheTrpLeuLysLysLysTrpLysSerThrIleAspLeuAlaValIleAsn 1282
 Qy 9976 ATCCCGAAGAGTAGTAGTAATGATGAGTTCGAAATCCAAATAGTATACATATAT 10035
 Db 1283 IleProLysSerAspTrpIleProThrLysLeuSerProAsnArgTrpIleProLys 1302
 Qy 10036 AGAAGTGTCCATATATGAAGCAAAATATATATATATATGAGAGAGATATCT----- 10086
 Db 1303 ThrSerGlyLysTrpArgGlyLysArgGlyLysIleTrpLeuGlnLysAspSerGlyTrpAsp 1322
 Qy 10087 AGTGGAGATGAGATTAATATATATGAGGACTTATCTCTCATATTTACT---TCATCC 10143
 Db 1323 SerGlyTrpTrpAspHisLys-----SerAspIleTrpSerSerSer 1336
 Qy 10144 GAAAGGAGATGAGATGAGATGATATATATATATATATATATATATATATATATATATAT 10203
 Db 1337 GlnSerGlyTrpGlnLysLeuAspIleAsnAspIleTrpAlaProArgAlaProLysTrp 1356
 Qy 10204 AAAACATTTGATAGAGTAGTACTAGAACCATCAAAAAGGATATATACCAACT-----GAT 10257
 Db 1357 LysThrLeuIleGlnValValLeuGlnProSerGlyAsnAsnThrThrAlaSerGlyAsn 1376
 Qy 10258 GATACCAACT-----AATGATACACCAGCTAGCAATATGATTT 10296
 Db 1377 AsnThrProSerAspTrpGlnAsnAspIleGlnAsnAspLysIleProSerSerLysIle 1396
 Qy 10297 ATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10356
 Db 1397 ThrAspAsnGlnTrpAsnThrLeuLysAspGlnPheIleSerGlnTrpLeuGlnSerGln 1416
 Qy 10357 GAACCAAT-----AATAATATCAAAAAGTGCAGATATTCATGCAATGATACGAACT 10407
 Db 1417 GlnProAsnAspValProAsnAspTrpSerSerGlyAspIleProLeuAsnThrGlnPro 1436
 Qy 10408 AATACCTTATATCTGATATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 10467
 Db 1437 AsnThrLeuTrpPheAspAsnProAspGlnLysProPheIleThrSerIleHisAspArg 1456
 Qy 10468 GATTATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10527
 Db 1457 AspLeuTrpSerGlyGlnGlnLysTrpSerTrpAsnValAsnMet---ValAsnThrAsnAsn 1475
 Qy 10528 GATATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10587
 Db 1476 AspIleProIleSerGlyLysAsnGlnLysTrpSerGlyLysAspLeuIleAsnAspSer 1495
 Qy 10588 CTATGTGTGCTAAACATGATGATATATATATATATATATATATATATATATATATAT 10628
 Db 1496 Leu--AsnSerAsnAsnValAspIleTrpAspGlnValLeu 1508
 RESULT 14
 T18399
 Variant-specific surface protein 1 homolog 3D/vari - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T18399
 R:Rubio, J.P.; Thompson, J.K.; Cowman, A.F.
 EMBL J. 15, 4069-4077, 1996
 A:Title: The var genes of Plasmodium falciparum are located in the subtelomeric region
 A:Reference number: Z18927; MID:96324444; PMID:8670911
 A:Accession: T18399
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2042 <RUB>
 A:Cross-references: EMBL:U53324; NID:g1297090; PID:g1297091; PIDN:AA09769.1

C:Genetics:
 A:Map position: 12
 A:Note: 3D/vari
 Alignment Scores:
 Pred. No.: 1,48e-68 Length: 2042
 Score: 1542.50 Matches: 697
 Percent Similarity: 27.58% Conservative: 280
 Best Local Similarity: 19.68% Mismatches: 657
 Query Match: 7.93% Indels: 1909
 DB: 2 Gaps: 110
 US-10-087-013-1 (1-10628) x T18399 (1-2042)
 Qy 496 CATGATTTATTTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 555
 Db 12 HisAsnLeuLeuLeuAspValCysLeuAlaIleGlnTrpIleGlnSerGln 31
 Qy 556 AATCAT---CCACATAAA-----GGAATTCGAGAGCTGTACTGCTTT 597
 Db 32 AsnHisGlyLysHisGlnLeuSerTrpProAspSerProSerGlnLeuCysTrpGlnLeu 51
 Qy 598 GCAGAGATTTGCGATATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657
 Db 52 AlaArgSerPheAlaAspIleGlyAspIleValArgGlyArgAspLeuTrpArgGlyAsn 71
 Qy 658 -----GTCCATGACAAAGTAGAAGAGCGGCTCCGAGAGCTTTTCAAGAAATACATGAT 711
 Db 72 AsnArgGlnAsnAspLysLeuGlnLysLysLeuLysGlyTrpTrpLeuLysLysIleTrpAsp 91
 Qy 712 GGAATG-----GAGATGAAAGTAAATAATGATTTACATTCCTGATGATGGA 759
 Db 92 AsnLeuValGlnLysLysLysGlnGlnLysLysLysLysLysLysLysLysLysLysLysLys 110
 Qy 760 AATTTATTTAATTAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 819
 Db 111 AspPheTrpIleLeuTrpGlnAspTrpTrpAlaLeuAsnArgGlnAspValTrpLysAla 130
 Qy 820 ATACATGTGATGCATCATATTAATCTGATATTTATGCAATGCAAGAGTAAATACACCA 879
 Db 131 IleThrCysAspAla---HisAspSerArgTrpArgLysMetGlyAlaAspGlySerIle 149
 Qy 880 TTATTTTCAATCTTAATATCGGCCATTAACAAAGAAAGTTCCTACCAATTTATGATTT 939
 Db 150 ThrGlnSerLeuMetArgGlnCysArgAsnValAlaAspValProThrAsnPheAspTrp 169
 Qy 940 GTCCCTCAATTTTACGTTGCTGCGAGATGAGGAGAGATGTTTCCGAAAAAGAAAT 999
 Db 170 ValProGlnTrpLeuArgTrpPheGlnGlnTrpAlaGlnAspPheCysArgLysArgLys 189
 Qy 1000 ATTAATTTGAAAAAGTCAAGACTCCTGCTGAATGAC-----AAGAAGCG 1047
 Db 190 HisLysLeuLysAspAlaIleGlnLysCysArgGlyClnAspGlyTrpGlnLysAspArg 209
 Qy 1048 TTATTTGATGATATATGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1107
 Db 210 ---TTCysAspLeuAsnAspGlyTrpAspCysLysArgTrpIleSerAlaLysHisGlnLeu 228
 Qy 1108 CATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1167
 Db 229 ValGlnGlnLysLysLysLysCysSerValValCysIleProPheGlyProTrpIle 248
 Qy 1168 GGAATTCACACAGACATTTAAAAAACAAGAAATATGAAAAAGAAATACAA--- 1224
 Db 249 AspAsnGlnLysGlnGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 268
 Qy 1225 -----TCATATTTATCGAAGATACAAATTTGTCATATATATATAT 1266
 Db 269 LysHisAspGlnTrpTrpLysGlnLysLeuSerGlyAsnArgTrpLysLysArgSerLeuThr 288
 Qy 1267 AGTCAATATATATATA-----CAATTTATGAAAAAGTAAAGAAAGCAATAT 1314
 Db 289 ThrLysAsnTrpLysGlyTrpAspGlnGlnPheTrpLysLysIlePheLysAsp---GlnTrp 307

QY 1315 GCACTAATGACACTTTTAAATTAATAAGAAAGATATTGTAAGGA----- 1368
Db 308 ProbspValaLysPheLeuAspLeuSerLysGluThrAlaCysGluSerGlnPro 327
QY 1369 ---GGATTACCGAGAGAAAGATTAATTAATTAAGTCTGATGACAAAGGATA 1425
Db 328 TyrAspGluProAlaGlyThrIleSerIleAsnPhelLysAsn---TyrLysAsnProAspIle 346
QY 1426 TTTTATGCTTGAATATGCGCAAGTGTCCCGACTGCGGGGCTC----- 1470
Db 347 PheSerHisThrGluTyrCysGlnAlaCysProIlePheCysGluMetThrMetTyrLeuLeu 366
QY 1471 -----AAATGATGATGATTAATAATACACACAAATACATGATGATGAT 1515
Db 367 MetValAsnValGlnLysThrGlnAlaGlyAsnTyrAlaThrIleLysLeuPheLysAsn 386
QY 1516 GAACGTGTAAATATGAAGACTATTAACCTCATGGGGTGTGAAGCTTCAATATACAT 1575
Db 387 ThrLeuIleGlnAsn-----ThrThrAspIlePro 396
QY 1576 GTCTTATATAGTGTATGACAAAGTATTAACAAATAATTAAGAAATTTTGTATC 1635
Db 397 IleLeuThrProAspThrThrLysSerAsnIleValGluLysTyrAlaGlyAsnPhelCysAsn 416
QY 1636 AGCTCAACTAATTAACAAGATTAATAATCAAAATGGGAATGCTATTAAGATGAA 1695
Db 417 SerSerAspAspAsnAsnSerAspGlnIleAsnAsnTyrPheLysHisTyrAsp----- 434
QY 1696 AATATTAATATGATGTAACATGAACAAATATCAAAATCAATATGATATAT 1746
Db 435 -----GluSerLysLysSerGluGlnAsnAspAsnLysValGlu 447
QY 1747 -----CCTAATATATCATTTTCAATATTTTGTGA 1779
Db 448 GltThrTyrPheGlnAsnPhelLysLysAspGlnLysValThrSerTyrAsnIlePheThr 467
QY 1780 TTATGGTTATCATATTAATTAAGGATTAATTAAGTATTAAGTATTAAGTATTAAGTAT 1839
Db 468 LysThrValSerGluMetLeuAspAspSerIleLysTyrAlaGluLeuAspLysCys 487
QY 1840 AATAAATAATACAAACCAACGATTTGATGAAATGTAACAAATGCTTATGTTTGCAC 1899
Db 488 LeuLysAsnAspLysLysThrCysGluLysLysAsnTyrLysAsnLysCysLeuAsnPhelLe 507
QY 1900 A---GATGGTTTAAACAAAGAAAGAAATGAATATGATTAAGAAATGCTTCAACAAA 1956
Db 507 LysAsnGluLysLeuAsnLysLysLys-----ArgAsnGluAlaGlyGlnLys 522
QY 1957 AAAAAGATATACAGCAATCGTATTAATTAATTAATTAATTAATTAATTAATTAATTA 2010
Db 522 GlyHisThrPheLysAsnLysLysIleAspAlaGlyLysLeuValCysLeuLeuLysTyrHisLe 542
QY 2011 -----TATTTTAAAGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2049
Db 542 LysValLeuLeuAsnAspValAlaPheLeuGlnAspMetGluLysAlaGlnLysAspProGlu 562
QY 2050 -----GCAAAATGAAAGAACTTAATTAATTAATTAATTAATTAATTAATTAATTA 2102
Db 562 HisIleLeuAlaLysIleLysGluLeuLeuLysLysAsnAspGluLysValAsnAsnLeuSe 582
QY 2103 CAATTTGGAATAATATAGGACTATTTAGACAAATGCAATAGCAACTCTTGTAGTACACTT 2162
Db 582 PheMetGluThrIlePheAspPhe-----LeuLeuGlnGluGlu 595
QY 2163 AAAAAGAACTGCAAGATATTAAGACATTAATTAACAAGAGAGATGTAAGATCCCA 2222
Db 595 uGluGlnAspAlaGlnLysCysValSerAsnAsnPro---GluLysCysGluGluThrGlu 614
QY 2223 TAATGCAACAACAACCCGCTGTGTTAAACCTCGGAGAGCAGCAACCACTAAATAATAT 2282
Db 614 nLysProIleThrAspGluAla-----ProGluGluAla----- 625

QY 2283 AAAAGAAATGACACAAATACCTTTAAAGAGACTGCATACGAGGAAGCAGAAATCGTGCTCT 2342
Db 625 ----- 625
QY 2343 TCATTAATTAAGAAAGAGCAGCAGAGATATATTAACGTTGGGGGTAGAGAAAGGA 2402
Db 625 ----- 625
QY 2403 CTTCAGAGCAATTTATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2462
Db 625 ----- 625
QY 2463 AAATGACCATGATGCGCAAGGACGAGTGATGATACAAACAGATTTGTCTAGC 2522
Db 626 ---GlyProSerProAspThrGlyThrAspAspAsnLeuGlu----- 638
QY 2523 AACTGAAATGGAGATGATGCGGACACATGCGTAAAGATCAAGAAATGTTATATATGCC 2582
Db 639 -----AspIleAspSerAsp---GlyGluGluAspAspAspVal----- 650
QY 2583 TCCTAGAGAGCAGACATATATGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2642
Db 651 -----SerHisValAspGluGluGluProGluAspAsnPro 662
QY 2643 ACTTAATGCTAATTAATTTGATGATTTAGTAAATTAATTAATTAATTAATTAATTAAT 2702
Db 662 oValGluGly----- 665
QY 2703 ATCAAGCAAAATTAAGAAACAAACAGATTAATCAATTAATTAAGAAAGAAATTAACCTAAA 2762
Db 666 ---SerSerGluGluGluLysGlnGluValValLys-----AspThrGluAlaAla 681
QY 2763 GGGCCCAAGAGATTAATGACCCCAACACCAACACTATCTGTGAGATATACGTTA 2822
Db 681 aValProLysGlnAspThrGlnProLysGluGluValAsnProCysLysIleValGluGlu 701
QY 2823 CAGTTTCCAGATTAATAGGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2882
Db 701 uLeuPhe----- 703
QY 2883 CATGTTAAAGCTGCAAGACATTTGGAACCTGTTTGGTAATTAATTAATTAATTAATTAAT 2942
Db 703 ----- 703
QY 2943 AGCGAAAGAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3002
Db 704 -----LysSerThrLysAs 708
QY 3003 TTGGTGGGAACCTAATAGACCCAAAGATGATGGAAGCCATGAATGTGATTAATAATTTT 3062
Db 708 nPheGluAspAla-----CysGluLeuLysTyrGlu 718
QY 3063 GAAGGATTAATCGGACACCAATCAACACAAAGTATTATTTGGGATATAGTATCATATC 3122
Db 718 LysAsnTyrGly----- 722
QY 3123 ACCATTGATGATTAATCCCAACAAATTAATTAATTAATTAATTAATTAATTAATTAAT 3182
Db 723 -----TrrLys 724
QY 3183 CTGCAAGGTCGAGAAAGAGATGATTAAGTTGAAGAGAAAGTGTAAAGAGTGAAGA 3242
Db 724 sCys----- 725
QY 3243 TAAGATTAATGTCGAAGGCTGTAGAAAGAGAGTGTAGAGTTGTACGAAGTGCACAGA 3302
Db 725 ----- 725
QY 3303 AGCTTGAATGAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3362
Db 725 ----- 725
QY 3363 AGATTAATTAACAAGATTAATCAATGACAAACACAAATATGCTGTAGTAATTAATTAATTA 3422

[illegible][illegible]


```
QY 5448 TGGTAGGATGATGAAATAGTGGAAATAGTGCAGAGATGAGATCTAAATAAAG 5507
Db 1029 ----- 1029
QY 5508 TGGTCTGTACCTTCAGATGATGATATTCCTATGGGAAAAATCCGATGAGACTGC 5567
Db 1029 ----- 1029
QY 5568 GTATCAGTTTCGTTCAGTGTTCGCCAATGGGTGAGATTTTTCGAAACATAAAGAAA 5627
Db 1030 ----- 1030
QY 5628 GGAATTGGAGAAATGAGAGGCGCTGT-----ATATATTACTGT--GGTATTA 5678
Db 1047 sargLeuAlaGlnIleTyrGluAspCysArgGlyAsnAspLysValCysSerGlyAspC1 1067
QY 5679 TGAAGAT-----AAAAGAAAGAAA----- 5697
Db 1067 yGluAspCysGluGluValArgGlyGlnAspTyrSerLysIleSerAsnPhenAsnCysPr 1087
QY 5698 ---TGTACAGATCGGTGTACACATATATAAAATTTATTATGAGTGAAGAACCCACAGTA 5753
Db 1087 oGlyCysGlyArgGluCysArgLysTyrLysAsnTrpIleLysThrLysArgThrGluPh 1107
QY 5754 TGAATAACAAATCAAAATAATGATGAGAAATTAAGACAAATAATATCCAGCATCTGT 5813
Db 1107 eAspGluGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1120
QY 5814 GGCAAAAGATGCAGAGAGACGCTCGCGAATTTTAGACAAAACATTAATAAAATTTGTGA 5873
Db 1120 ----- 1120
QY 5874 AAATTAAGATGAGATTTGTGAATATATAGTGAAGATGTGTCCACACAGCATTAAC 5933
Db 1120 ----- 1120
QY 5934 TGATGTATATAGTCAAAATATGCCCGCATGATTAGACATGAACAAAGAAAGTTGAAG 5993
Db 1120 ----- 1120
QY 5994 AAAGTAAATTTGTCAAGTGTCCAGAGGTCCACACAGTGTACAGAGGAAACACCGTCACC 6053
Db 1120 ----- 1120
QY 6054 ACGGATATCATGTATTCAAAGACGACGATCGAAAAAGAAAGCAAGCGCGCGCC 6113
Db 1120 ----- 1120
QY 6114 TACAAACAGCGCGAAAAAGTGAATAATCTAACACAGAAATGCGACACAAACAGCAAGC 6173
Db 1120 ----- 1120
QY 6174 CCGACGAGCAGACAAACAAACAGAAAGCAATCAACAGCAACAGCAAGCAATCTGA 6233
Db 1120 ----- 1120
QY 6234 CGTGGCACAATGTAAAGCCATCTTTCGAAATTAACAGATACAGGGGTGGAATAGA 6293
Db 1120 ----- 1120
QY 6294 GGGTGTATTCMAAAAGATATGACAAATATCTTAATGGGGTGTATTTAGGTAAATGC 6353
Db 1120 ----- 1120
QY 6354 TAAAGAAAATGAAATGGCATATGTATCTCTCTAGAGAAAAAATATATGTATTAATA 6413
Db 1120 ----- 1120
QY 6414 TATACAAATTTAATATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 6473
Db 1121 ----- 1126
QY 1121 ----- 1126
QY 6474 TATTAATGTGCAGCAATAGCAAACTCAATTTTGTGGTTAAATATATATATGAATCC 6533
Db 1126 ----- 1126
QY 6534 TGCAGCAAAATGAATTCGAAATATGCAAAATGCAACATTCAGATGAATTTAAAGATATGA 6593
Db 1127 -GlyAlaGluSerLys----- 1131
QY 6594 TTATACATATGTCAATTAATGAATATGTTTTTGGAACTGATATTTCTAATGATAAAA 6653
Db 1131 ----- 1131
QY 6654 AATTATACGTACAAATATAGTGAACCAACATTCATGAAAAATTAAGAAAAACA 6713
Db 1131 ----- 1131
QY 6714 GATTAATAAATAAGATGAGAAATTAAGTAAATATTTGGGAAAAATTAATAATTTAT 6773
Db 1131 ----- 1131
QY 6774 TTGGAGAGATGATATATGATTAACCTTATCATCTACAGACGAAAAAGAAAAAGAAA 6833
Db 1132 ----- 1133
QY 6834 AATTAGATATATATACAGTACATGACATGACCAAACTGACGCTTCCCTTGAAGATT 6893
Db 1133 ----- 1133
QY 6894 TGTAAAAAGCCCAATTTTGTAGATGTTGCACAGAAATGGCAGAGAAATTTGTAAATA 6953
Db 1133 ----- 1133
QY 6954 GAGGAAGACAGTGTAAATTTGAGCGCGGCTGTAAGAAATATGATGATAGTAG 7013
Db 1133 ----- 1133
QY 7014 TATAGCGTAAAGACACAGAAATGTGCAGAGCGTGTATACATATCAAAATTTATTA 7073
Db 1134 ----- 1142
QY 7074 GAATGGAATACTGATATATGAAGACAAAGCAAAAGTCAAAAAGATTAAGATGCGCA 7133
Db 1142 sLysLeuGlyThr----- 1146
QY 7134 AAATATTAAGATATCTCTCTACGAAGACATGAGAGAG---GCAACATGTCTCA 7190
Db 1147 ----- 1163
QY 7191 TGAATATTTAACTGAATTAATAAGCAATTAATGTGCATTAAGATTTCTTGTATGCA 7250
Db 1163 s-----AsnAsnLysAsp----- 1167
QY 7251 AAAACCTTCTCAACTACCAAAACACACAAACATCAATCAATCGCATGCTAATGA 7310
Db 1168 ----- 1171
QY 7311 TATGCCAATTCGCTGATATATGTTCTGTAAGATTAACAAGTGTGAGTGTCTGAAGT 7370
Db 1171 ----- 1171
QY 7371 TTCAAAAAGGATCTATGATTCATACAAAAAATTAAGTAACTTAATACCTATGAA 7430
Db 1172 ----- 1178
QY 7431 TTGTGTAGAGAAAGCAGCATATTAATTTATCTAAGACGAGAAATTAATGATATATAC 7490
Db 1179 ----- 1181
QY 7491 CTTGAAGAAAAATTTATATCATATGAGTCTACAAAGAAAAAGAAATTAATATGTTG 7550
Db 1181 rPheLys----- 1183
QY 7551 GACTAATATTAATCTTGGATCTTAAGAAACCTTATGACCTGTATTAATATAGAGAG 7610
```

Db 1184 -----ProAlaaspasncyslys----- 1189
QY 7611 AGAAMCCCTTGTGAATAAGAGAAATCGTTTAAAGTAGATATGATGAATGGAATG 7670
Db 1190 -----ProCys-----SerGluPheLysVal-----LysCys 1198
QY 7671 TTACAAAATTCAAAGTTCTATCAGAGAAAAAGAGTA---TGCTACCTCCAGAG 7727
Db 1198 s---GluAsnGlyLysCysSerGlyGlyLysValAsnCys----- 1212
QY 7728 AGAACATATGCTTAGAATTTAGATGAATTTAAATGAAAGACTTAGATAGTA 7787
Db 1212 ----- 1212
QY 7788 TTATTCCTAAATAATGCTGCAACTGCAGAAATGAAGAAATACATTAATAAAAA 7847
Db 1213 -----AsnArgLysAsnThrIleAlaIleAlaThrGluIleAlaAsnMe 1226
QY 7848 CTTCACACTCAGAGACGGGTGCGCAATGCAATATGTAATGTAATGTAATGTAATGTT 7907
Db 1226 tIleAsnSerThrGluGlyValPheMetThr---ValSerAspAsnSerAspHisLysPh 1245
QY 7908 CGCTGATCTGGTGACATCTTAGAGACAGATATGTAAGATTTGCGTTACTTACC 7967
Db 1245 eGlu-----GlyLysLeuGluPr 1251
QY 7968 TCCCGTAGAATAAATATATATAGCTTTTGAATACATATATGAAAAAGCAAAATAA 8027
Db 1251 ocysGlySerAlaAsnIlePheLys----- 1259
QY 8028 AAATAAGTAGAATAATATACACAGATGACAAACGTTTCGTTCTGTTGGGATGC 8087
Db 1260 -----Glu 1260
QY 8088 TAATAGAAAATATTTGAAGCAATGACGTGCAAGACACAGAAATGCAAACTTTT 8147
Db 1260 yIleArgLysAspValThrLys-----CysGlyGluThrCysGlyValAspIleCys 1277
QY 8148 TAGAAAAGAGAGATGATTTGAACGATTAACATTAATACAGATTAAGTGTGACA 8207
Db 1277 sGluProAsnThrPheAspGlyLysGluAsn----- 1287
QY 8208 TAAAGACATCCACTGTATGATATATATACCTCAACGTTTGCATGATGATGTAATG 8267
Db 1288 -----GlyLysGluThrIleGluIleArg---AlaLeuLeuLysArgTr 1301
QY 8268 GTCTGATATATTTGAAGCAGCTGATGAGATTTGGAATAATTTAAAAATCATGTGA 8327
Db 1301 pValGluThrPhe-----LeuGluAspTyrAsnLysIleLysHisLysIleSe 1317
QY 8328 TCACTGTAACATCTGACAGA-----TCCAGAGATGATATGATGAAATAAGTGA 8378
Db 1317 rHisCysMetArgLysGlyGluLysThrIleCysIleAsnAsp----- 1331
QY 8379 TGAACAGTGTAAACGAGATGTCAAGATATTAATAATTTGTTCTTAATGCAATCTCT 8438
Db 1332 -----CysValGlu-----LysTrp 1336
QY 8439 ATTGCAATATCAATCAATAATAATCAAGAAATTTGATGAACACCAATATATCAAAAT 8498
Db 1337 -IleAsnIleLysLysLysAspTrpGluThrIleArgGluArg---TyrValLysGlu 1354
QY 8499 CTCTACTATGATCATGTTCAAAATTTTGTACAAAAGTTGAAAACCTTTAAAGTGAATG 8558
Db 1354 nTyrThrThrGlyHis----- 1359
QY 8559 TTCTGTGAGAGCTTTTCTGAATATCTTCATGAACAAGTAAGTTTGAATTAATTT 8618
Db 1359 ----- 1359
QY 8619 TAATGAATAATGATGTTCTTCCATATATACGAACATATGCTTTCGAGAAAACACCAAAAG 8678
Db 1360 -----SerAspIleTyrLysValThrSerPheLeuLysProGluInP 1374
QY 8679 TTATTAAGACCTTGACAGTTGTACACTCTTCAAGAAATCCATTTGATATGCTCTAC 8738
Db 1374 eHisAsnGluValLeuLysAlaIleLysPro----- 1384
QY 8739 CGATCAAAACAAAGATGAGTAAAGATTTCAAACTTTTACCTTCTGCTCGAAGAAATGA 8798
Db 1385 -----CysGlyAspLeuAspLysPheGluIn----- 1392
QY 8799 TTATGATTAATATCTTGATTAATGGAACGATACCTGTTCTTAATATGATCGATGATA 8858
Db 1393 -----AsnSerThrAsp----- 1396
QY 8859 CAAAGGTATATGATTCCTCCAGAGAGACATTTATGTACAGACCTATCATCGATA 8918
Db 1397 -----CysThr----- 1398
QY 8919 TAATTAAGAAAAGTGATTAAGAAATTTAAAAAAAACCTTCTACTCTGCTTTCAG 8978
Db 1399 -ValValArgSerSerGluAsnGlyValThrAsnLysLys----- 1411
QY 8979 TCAAGACAAATTTGTAGTCAAAAATTAATGAAGAAAGATTTGCTTGAAGCAAT 9038
Db 1411 ----- 1411
QY 9039 GAAATATATGTTATGACATTAATTCAGATTAATTAAGAAACATGATATGATGACACTTC 9098
Db 1412 -----AspIleValGluCysLe 1417
QY 9099 ATTATCTGAAAATAATTAATAAATTTTGAACAATCAAAATGACAAACGAAATGCTAA 9158
Db 1417 uLeu-----GluAsnLeuLys 1422
QY 9159 AACATGCTGGAAAATAATAGACGTCAATATGCGACGCTATGTTATGATGATTAATAAT 9218
Db 1422 sThr-----LysAlaLysThrCys----- 1428
QY 9219 TGTACTTCAAAAGTAACATTAATGATGAAGATGCTGTCAATTAACAAAGATGAAGAAC 9278
Db 1429 -----Pr 1429
QY 9279 TAATCAGTTTCTGTTGTTAATTTGAATGGGCAACACATGTAAGAAAAACA 9338
Db 1429 oAsnGlu-----ValAsnGlyLysAsnGluThrCysAspSerLeuProH 1444
QY 9339 TGTAGTGAATCATTAATAACAAATATGCTCGTCAACAGAGATATTTGAAGCGTC 9398
Db 1444 sValGluAspAsp-----AspAspGluGluProLeuGluGluIn 1457
QY 9399 AGAATTAATTAAGACACACCTGATGCTCAGAAATGATATTAAGAAATATATGCTTGAATAT 9458
Db 1457 rGlu----- 1458
QY 9459 ATTGATTAATAATCAATGCAAAATCTTAATATTAATAATATAGCAATTAAGATCATC 9518
Db 1459 -----GluAsnThrValGluGluInProAsnIleCysProGluInLeuProLysProProPr 1476
QY 9519 TTCAGTAATATAGACATATTAACCATCTGAGAGAAAGAAAGTTGATGATATTAATAATCAAA 9578
Db 1476 o-----GlnProGlyAspGluAspGlyCysLysGluAlaSerProAl 1490
QY 9579 AGATTCGAATCGCTTTCGAGTTAAATGATATTAATGAATTAAGTATACAGAGACAAAATA 9638
Db 1490 aProSerGlu-----GlyThrGluAs 1497
QY 9639 TAATGAATAATATGAATTTCAAGAAGATTAATAAATAATTAATCTGTTATATTTTGT 9658
Db 1497 nGlu-----ProProVal----- 1501
QY 9699 TGAAGATGAACACACAAAATAATCATGTACTGATGAAATTAATAAGACAGACAA-- 9756
Db 1502 -----IleLysProGluGluGluAl 1508

Db 768 -----AlaThrGluGluValAspAla 774
 QY 2935 TCACGCCAAAGGAAAGAAATGATAAATATGATGATGCCCAATATTAAATATG 2994
 Db 775 AsnAlaSerSerAspGlyGluAspAspGluGluGlu ----- 787
 QY 2995 AGGGAATAATGGTGGGAACCTAATAGAGCAAAAGTATGGAGCATGAAATGTGATATA 3054
 Db 787 ----- 787
 QY 3055 AATATATTGAAGATAAATCGGAGACCAATACACAAAGTAGTTATTCGGATATAGT 3114
 Db 787 ----- 787
 QY 3115 GATCATACACATGGATGATATATATCCCAAAAATTAAGATGATACCGAATGGCA 3174
 Db 787 ----- 787
 QY 3175 GAATGTACTGCAGAGTGCGAAGAAAGAGTATGATAGTTGAGAGAACTGTAGAGAG 3234
 Db 788 -----Glu 788
 QY 3235 TGTAAAGATAAGATAATAGCTCAGGCTGTACGAAAGAGAGTGCTACAGTTGTACGAAG 3294
 Db 789 GLuGluGluGluAspGluGluGlu ----- 796
 QY 3295 TGCACAGAAGCTGTATGATATGATATATGATATATAGATTTATGAAAGACAAATGGAAT 3354
 Db 796 ----- 796
 QY 3355 ATATATCAGATTAATACAAAGATTAACATGACAAAGCAACAAATGTCTGTATGATAGT 3414
 Db 797 -----GluGluAlaGlu ----- 800
 QY 3415 GGTATGAAAGCTTCCAGTACTGCCAAAATCATATAGACAGAAATGTATGATTTTGG 3474
 Db 800 ----- 800
 QY 3475 TCGAATTTATACCAAAATAGTGGCGAAGATAAAGATGCTAGTATGATGAAGT 3534
 Db 800 ----- 800
 QY 3535 GCTGTCAATGGTACTAACACACAGTATGAAATGTTGGACATATCTCCATGATACAGGA 3594
 Db 800 ----- 800
 QY 3595 AATTTGATGATTTGTCAGTACAAAATAGTTTGTGATGAAAAAGTGTGTAAGAT 3654
 Db 801 -----GluValGluGluGluThrAspGluSerAla 811
 QY 3655 AACGAAATAATAGCTTTAGATTAACACAGAGCAATGATGCTGCTGTGTA 3714
 Db 812 ThrGluValAla ----- 816
 QY 3715 AGTGGATCGAAAGCAGACAGGTACATATAAAGAAAAAGGAGAAAGGAT 3774
 Db 817 ---ProSerProGlyThrThrGlnAspGlyValIlySerProAlaSerGlnGluAsp 835
 QY 3775 ACGGAA---GTAAGACAGTGAATGATGATTAAGAAAAAGCAATGATGAAAGAAACAAGTA 3831
 Db 836 ValIlyValIlySerIleValAspIlyAlaLeuIly -----GlyIlySerLeuAsp 851
 QY 3832 GAAGATGTGATCCAAAAAAGAAATAGTAATGATATCCGATTTGGCAATCGGAAATATA 3891
 Db 852 AspAlaCysThrIleuIlySerIlyGlyThrAlaProThrSerTrpLysCys ----- 868
 QY 3892 AATTTAGTGAAGACCTCGI----- 3912
 Db 869 ---IleProSerAspThrLysSerValAlaThrThrGlySerAspThrThrGlySerGly 887
 QY 3913 ---GTGTATAGCCCTAGAAAGCAAAAGTTATGCTGATTTCTTG----- 3957
 Db 888 SerIleCysValProProAlaArgArgIlyLeuIlyValGlyIlySerLeuHisAspTrpAla 907

QY 3957 ----- 3957
 Db 908 GlyGlyGluThrThrGluAlaLysSerGlnGluThrSerGlyGlyGlnIlyThrProSer 927
 QY 3958 GCAAATGATAAT-----GAAATAAAAA 3981
 Db 928 GlyAsnGluSerSerProSerGluIlyLeuProGlnIlyProThrProGluThrLys 947
 QY 3982 TTACATACCAAGTTAATTTAAAGAGCTTTCATCAATCTGCAGACAGCAAGAAATTC 4041
 Db 948 GluThrProGluSerSerLeuLeuHisAlaIleValSerProProAlaGlyLeuArgPhe 967
 QY 4042 TTCATGTGATATTAATAAGTAAAG----- 4068
 Db 968 LeuProThrIlySerPheIlyGluGlnIlyThrPlySalGlnIlyAlaGlyAlaThrGly 987
 QY 4069 -----GATGTGAAGAAATGAATCGCAATGAATATAA 4104
 Db 988 LeuGlnLeuProGlyValThrValAspAspSerAspProAspProGlnIlyThrLys 1007
 QY 4105 GAAGCAAAATTCCTCCCGCATTTTGTAGATCCATGTTCTACACATTTGAGATTATAGA 4164
 Db 1008 ArgGlyAsnIleProAsnAspPheLeuArgGlnMetPheIlyThrLeuGlyAspTrpArg 1027
 QY 4165 GATTTTATTTGGAACAGT-----ATATCAAAA 4194
 Db 1028 AspIleCysIleGlyGlyAspArgAspIleValGlyAspThrIleValSerIleThrGlu 1047
 QY 4195 GGTCAATGTGAGGAGAAATTAATAAGCAAAAGATGCTCTTTCAAAAAGTGTCAC 4254
 Db 1048 GlyIlySerThrLysLysIlySerIlyIleIleGluGlyPheLeuLysGlnIlyThr 1067
 QY 4255 CAAAAATCTCTAATGAAAA----- 4275
 Db 1068 ValThrSerProSerProArgAspThrSerSerArgThrProValIlyProGlnIlyThrSer 1087
 QY 4276 -----ACACGCCAGATGCTGGACAGAACATAGTCATGATATGGAAGCT 4323
 Db 1088 ValGluLysThrProGlnIlyThrThrProGluAlaAsnGlyProHisIleThrAsnGly 1107
 QY 4324 ATGCTATGTGACATGTA----- 4341
 Db 1108 MetIleCysAlaLeuThrTyThrGluAspSerGlyAlaIleGlyGlnProGlnIlyVal 1127
 QY 4342 -----AAAAAT-----GGGCAAAA----- 4356
 Db 1128 GluAspAlaAspLysValLeuGluIlyLysLeuLysProAsnThrAlaAsnGlyIleLysTrp 1147
 QY 4357 -----AAGATGATTTTTCGAAAACATACGCTTACAAACACGTCAAATTTAGTGACAAA 4410
 Db 1148 TyThrLeuLysGluAspAsnThrSerSerAlaMetProThrSerSerSerSerGly 1167
 QY 4411 AGCACC-----ACTTGAAGATTTGCCAAAGCAAGCCAGTTTGA 4452
 Db 1168 SerAsnAspProIleAsnThrProLysLeuThrGlnPheValGluIleProThrPhe 1187
 QY 4453 CGATGCTTACCGAATGTCACACACATGTTCTATACACGCAAAATTTGAAAGAT 4512
 Db 1188 ArgTyThrLeuHisGluThrProIlyGlnAspPheCysLysGluArgMetLysArgLeuGln 1207
 QY 4513 GTGCAGAAAATGTAAAGCAATGAC----- 4539
 Db 1208 IleTyThrGluGluLysValGlyGluAsnGlyTyThrGlyArgGlyArgGlyGlnIlyThr 1227
 QY 4540 ---CAATGAAGT-----GAT 4554
 Db 1228 ProGlnCysSerCysTyThrGluAspCysGluAspGlnLeuSerLysTySerTyAsp 1247
 QY 4555 ACA-----GAATGATTAAGAAATGCGAGACATGCTAAATAT 4593
 Db 1248 ThrValAlaAspLeuGluCysProLysCysAlaLysHisCysArgTrpTyThrLysLysTrp 1267

```

QY 4594 ATGAAAAA-----GAGTGGATTCCACAGATAA----- 4629
Db 1268 TLeGluLysLysAspGluPheThrGluGluLysAlaPheProLysGluLysAsp 1287
QY 4630 TATTACAGAGTACGACGACAAAAAGATTGATAGA----- 4668
Db 1288 TyrTyrValAsnGluYasnAsnLysGlyGlyGluAspAsnGlyPheCysIleThrLeuLys 1307
QY 4669 -----CAACACATTGGCTTAATGCTTACAGACTATACT 4701
Db 1308 SerLeuSerAspAlaIleGluPheLeuGluLysLeuGlySerCysLysAspAsnSer 1327
QY 4702 GGAACGATGACACAGATTACTTG-----AACAGAAATTACTGCTAGT 4746
Db 1328 GluAspAsnGluYasnAspLysLeuAsnPheSerGlnProAsnGluThrPheValProAla 1347
QY 4747 TGTGTGATTAAGCTCGACGCTGCTGTGCTACAAAGAAATATACATTGTTCAGAAAA 4806
Db 1348 ThrAsnGlySerProCysSerGluPheLysIleAspCysLysGluAsnGlyLysCysLys 1367
QY 4807 CAGGCTTACTATGATGCCGACAAACATTGGGTGCGACAAATTTATGAAAAATGACGAC 4866
Db 1368 AsnGlyGlyGlyGlyThrAsnGluThrCysAsnGlyThrThrPheIle----- 1383
QY 4867 AATATATCTAACATTCGAGTAAGAATTAAGTCAAAAGGATTAGTAAGAAGCGAACACA 4926
Db 1384 -----ThSerGluAsnPheLysGluLysGly-----GlnThrAlaLysGlu 1397
QY 4927 GGTCTATTAAGTGGCAAAACAAAGCTCTAATTAATCAATTAATCAATTAATCAATTAAT 4986
Db 1398 PheValMetArgValIleSerAspAsnProAsnGlyPheAspAspLeuAsnGluAlaCys 1417
QY 4987 GAAGATGTC-----CTTTTCTTCTCTCGACATA-----CGTAAATGT 5025
Db 1418 GluAsnAlaGlyIlePhePheSerIleArgLysAspGluThrGluCysGlyLysValCys 1437
QY 5026 TTTATGATGATG-----GATGGCAATTATACAGATCCAGCAAGTTAA 5067
Db 1438 GlyTyrGluValCysIleProGluLysGlyAsnGlyValIleThrThrSerGlyGluAsnAsn 1457
QY 5068 GATGAA-----AATGGTTGCGAAAAAGATTGATGGAAGTGGCGCAACGAA 5115
Db 1458 AspGlnIleIleThrIleArgLysLeuValAlaIleThrValGln---AsnPheLeuAsp 1476
QY 5116 GGGTACAAATTGGGTCAATCTACAAAGAAAAAGAAAAAGAAAAATTAATAAACGTCG 5175
Db 1477 AspTyrAsn-----LysIleLysHisLys 1484
QY 5176 GATGGCGAC-----AATATTTCTATGAGTCCGCGCTTGT----- 5211
Db 1485 IleSerHisCysLysAsnSerSerGluGlyTyrThrCysIleLysAsnCysValGluGln 1504
QY 5212 -----AGTCCATTAAGAAATTAAGTTTATGATTTAAGATATTAATTTAGATTGAT 5265
Db 1505 TrpIleSerThrLysArgThrGluTrpThrAsnIleLys---IleLeuLeuAsn----- 1521
QY 5266 AATTGGAGATGAAAAAAGACCGAGAAAAATTTGAAGAAATA----- 5313
Db 1522 -----GluGlnTyrLysAspAsnProAspTyrAsnValLysThrIleLeuGlnAspLeu 1539
QY 5314 -----TTTACAAAAAT-----GGAACATGATGTCGCAAGA 5346
Db 1540 GlnSerGlnIleAspPheAsnIleAlaIleLysProCysGlyThr---LeuThrLysPhe 1558
QY 5347 AGTATAGTACTACAGAAATCCCGATAGTACTCCGCAAAATTTTCTGAAACGAATAAT 5406
Db 1559 GluAspSerCysGlyIleAsnGlyAlaGluSerSerGluLysLysAsnGlyHisGlu--- 1577
QY 5407 AAGCAATGTGTGGAAACCAATGATGCGGGTACAAACGTGTGACG-----GAT 5457
Db 1578 -----TyrAspAlaIleAspCysMetLeuAsnArgLeuGlnAspLysIleAsp 1593
QY 5458 GATGAAATGATGTAATTAATGCAAGAAAGTATGAAAGATCTAAAAAATGT-----GAT 5511

```

```

Db 1594 AspCysHisLysAsnHisIleAlaGlnAsnGlyGlyGluAsnGlnAlaLysCysGluLysHis 1613
QY 5512 TCTGTACCTTCAGATGATGATTAATCTATGCGGAATAATGCGATGAAGTACTGCGTAT 5571
Db 1614 SerAlaProAspGluAspAspAspGluAlaIleGluGluGluAsnProValThrGln--- 1632
QY 5572 CAGTTTCTTGATGCTGTTGCCGAATGGGTGAGATTTTTC-----AACATTAAGAAAG 5628
Db 1633 -----ProAsnIleCysProLysProProGluPro 1642
QY 5629 GAATGGAATAATTCGTAAGGCGGTGAATGATTAATCTGTGTGTATTAATGAAGATAAA 5668
Db 1643 LysAlaGluGluLysGlyGly-----CysGluProAlaGluLysLys 1656
QY 5689 AGAAAGAAATGTACAGATCGGTGACACAAATATTAATAATTTATGATGATGGAACCA 5748
Db 1657 GluLysVal-----GluGlu 1661
QY 5749 CAGTATGAAAAACAAATCAAAAAATATGCTGAGAAATTAAGCAAAATATATTCGAGCAT 5808
Db 1662 LysGluGluLysThrValAsnThrValAlaLysProThrGluLysGluAlaIleGlyAsp 1681
QY 5809 CCTGTG-----GCAAAAGATGACAGAGACGCTCGAATATTTAGACAAACAAATTA 5862
Db 1682 ProAlaGlyProAlaIleAspSerGluGluAsnProGlu----- 1694
QY 5863 AAAATTTGTAAAAATAAAGTGGAGATTGCAATATTAATGTAAGATGTAAGATGTCGACA 5922
Db 1695 -----GluGluLysAlaProGluProGluValGluThrLysLysAspLysAlaPro 1711
QY 5923 CAGGATTAATCTGATGTAATAGTCAAAATATATGCCCCCATCA 5964
Db 1712 ValLysProThr-----ProIleSer 1718

```

Search completed: April 28, 2003, 12:08:49
 Job time : 1240 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

SUMMARIES

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 28, 2003, 10:52:36 ; Search time 418 Seconds
(without alignments)
6776.017 Million cell updates/sec

Title: US-10-087-013-1

Perfect score: 19455
Sequence: 1 atgggggtctctcgaataa.....atatatgatgaagatattg 10628

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

rched: 908470 segs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ n2p.model -DEV=x1p
-O/Cpu2.1/USPTO.spool/US10087013/runat.28042003.102844.20247/app.query.fasta_1.10823
-DB-A.Geneseq.101002 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -DOOPT=0
-LOOPT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdl
-LIST=45 -DOCALL=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US10087013 @CN1_1.721 @runat.28042003.102844.20247 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database: A_Geneseq.101002.*

1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	19407	99.8	3542	22 AAB62142	P. falciparum FCR3
2	3501	18.0	3060	18 AAM22475	Plasmodium var-7
3	3501	18.0	3060	21 AAV77905	Plasmodium var-7 p
4	3128.5	16.1	2913	17 AAM00384	Plasmodium falcipa
5	2972.5	15.3	2710	18 AAM22482	Plasmodium Proj3.
6	2972.5	15.3	2710	18 AAV77904	P. falciparum Proj
7	2972	15.3	2703	16 AAR70236	P. falciparum Proj
8	1773	9.1	2197	21 AAB18352	Plasmodium falcipa
9	1755	9.0	2182	18 AAM22476	Plasmodium var-1 p
10	1723.5	8.9	2228	20 AAM93944	Plasmodium pfm
11	1244.5	6.4	1726	17 AAM00385	Truncated plasmodi
12	1064.5	5.5	1700	21 AAB18144	Plasmodium falcipa
13	1032.5	5.3	700	16 AAR70235	P. falciparum EBL-
14	1032.5	5.3	700	16 AAM22481	Plasmodium ebl-2.
15	1032.5	5.3	700	21 AAV77903	P. falciparum ebl-
16	1032.5	4.3	431	21 AAB18350	Plasmodium falcipa
17	835.5	4.3	407	22 AAB62148	P. falciparum varc
18	827	4.3	1086	23 AAV76760	Plasmodium falcipa
19	665.5	3.4	294	22 AAB62147	P. falciparum varc
20	655.5	3.3	921	18 AAM22480	Plasmodium E3la.
21	649.5	3.3	921	21 AAV77902	P. falciparum ebl-
22	649.5	3.3	1604	16 AAR70105	TNF-R-EBA 175 fusi
23	636.5	3.3	1786	14 AAR41043	CD4-EBAI75 fusion
24	632.5	3.2	793	16 AAR70234	P. falciparum E3la
25	631.5	3.2	1435	16 AAM22477	P. falciparum SABP
26	618	3.2	1435	16 AAM22477	Sialic acid bindin
27	618	3.2	1435	16 AAV77900	P. falciparum SABP
28	609	3.1	311	22 AAB62150	P. falciparum varc
29	607	3.1	1501	23 AAV76762	Plasmodium falcipa
30	607	3.1	1421	23 AAV76764	Plasmodium falcipa
31	600.5	3.1	1143	23 AAV76759	Plasmodium falcipa
32	597	2.9	308	22 AAB62151	P. falciparum varc
33	572	2.8	351	22 AAB62149	P. falciparum varc
34	550.5	2.8	440	21 AAB18146	Plasmodium falcipa
35	539.5	2.7	616	23 AAM50533	Unidentified amino
36	529	2.5	445	22 AAB65344	Malarial parasite
37	495.5	2.5	3973	21 AAB18253	Plasmodium falcipa
38	453.5	2.3	749	16 AAR70233	P. falciparum EBL-
39	431	2.2	749	18 AAM22479	Plasmodium ebl-1.
40	431	2.2	749	21 AAV77901	P. falciparum ebl-
41	431	2.2	1979	23 AAB18171	Plasmodium falcipa
42	427	2.1	10182	23 AAB38314	Staphylococcus epi
43	418	2.1	248	21 AAB18151	Plasmodium falcipa
44	409	2.1	5024	22 AAG82935	S. epidermidis ope
45	408.5	2.1	5024	22 AAG82935	

ALIGNMENTS

RESULT 1
ID AAB62142
AAB62142 standard; Protein; 3542 AA.

AC AAB62142;

DT 29-MAY-2001 (first entry)

DE P. falciparum FCR3 varCSA protein.

XX FCR3 varCSA protein; chondroitin sulfate A; CSA; var gene; PREMP1;

KW erythrocyte membrane protein 1; parasitized red blood cell; FRBC;

KW malaria; protozoacide.

XX Plasmodium falciparum.

OS
PN WO200116326-A2.
XX
PD 08-MAR-2001.

QY 1621 GAAATTTTGTACACCTCACTAATTAACAAAGATAAAAATATCAAAAAATGGAATGC 1680
|||||
Db 541 GluAsnPhcysAsnSerThrAsnTyrLysAspLysAsnAsnGlnLysTrpGlnCys 560
QY 1681 TATTATAGAGTGAATAATATTAATAGTGTAACTGAGACAAAATCTGAAATCAATAT 1740
|||||
Db 561 TTTTCTTysAspGlnAsnIleAsnArgCysLysLeuGlnGlnAsnThrGlnIleAsnAsn 580
QY 1741 GATATCTCAAGATATATATCATTTCAATTTTTCGAATTAATGAGGTATACATATTTATTA 1800
|||||
Db 581 AspAsnProLysIleIleSerPheHisAsnPheGluLeuThrValThrTyrLeuLeu 600
QY 1801 AGGCACTATTAAGTGAAGCAACAACTTAAACTGTATTAATATATACACACGAT 1860
|||||
Db 601 ArgAspThrIleLysTrpAsnAspLysLeuLysThrCysIleAsnAsnThrThrHis 620
QY 1861 TGATATGATAGTAAACACAAAATGCTTTGTTTGACGATGGGTAAACAAAAGAA 1920
|||||
Db 621 CysIleAspGlnCysAsnArgAsnCysLeuCysPheAspArgTrpValLysGlnLysGln 640
QY 1921 GAAGCAATGCAATAGTAAAGAACTGTTCACAAAAAAGAAATATACAGCAATCGTAT 1980
|||||
Db 641 GlnGlnTrpAsnSerIleLysLysLeuPheThrLysLysLysAsnIleGlnGlnSerTyr 660
QY 1981 TATATGATATTTAATAATCTTTTGAAGGTATTTTAAAGTTATGATTAACCTTGAC 2040
|||||
Db 661 TyrSerAsnIleAsnAsnLeuPheGlnGlnTyrPhePheLysValIleAspLysLeuAsp 680
QY 2041 AAAGATCAACCAAAAATGGAAGAACTTATGGAATAATTAAGAAAATGAGTTT 2100
|||||
Db 681 LysAspGlnLysLysTrpLysGlnLeuMeGlnAsnIleLysArgLysLysAsnGlnLysPhe 700
QY 2101 TCCAAATTTGAAAAATATAGGACTATTTAGAGATCAATAGCAACTGTATAGATCAC 2160
|||||
Db 701 SerAsnLeuLysAsnAsnArgAspTyrLeuGlnAsnIleIleGlnLeuLeuLeuAspHis 720
QY 2161 TTTAAAGAACTGCCAGATATGTAAAGACATTAATACAAACGAAAGCATGTGAACATCC 2220
|||||
Db 721 LeuLysGlnThrAlaThrIleCysLysAspAsnAsnThrAsnGlnLysGlnThrSer 740
QY 2221 CATATATCAACAAACAAACCCGTGTTAAACCTGCTGAGGACGCAACCCACTAAATAT 2280
|||||
Db 741 HisAsnAlaThrThrAsnProCysValLysProArgLysLysThrGlnProThrLysAsn 760
QY 2281 ATAAAGAAATAGCACATCTTTAAAGAGATGCATACAGGAAACGCAAAATCTGGT 2340
|||||
Db 761 IleLysGlnIleAlaGlnTyrPheLysArgSerAlaTyrIleGlnIleAlaArgAsnArgGly 780
QY 2341 CTTCATATAATTGAAGAAAGCAACGAAAGTATATATTAACGCTGGGGTAGAGAAAG 2400
|||||
Db 781 LeuHisLysLeuLysGlnLysAlaHisGlnGlyIleTyrLysArgGlnLysArgArgLys 800
QY 2401 GACTTCAAGGACAATTAATAGATTAATATGAATAAATCATCTAATCTGATCTTGTTT 2460
|||||
Db 801 AspPheLysAspHisLeuCysArgIleMeCileLysHisSerAsnArgAsnLeuGlnLysPhe 820
QY 2461 TCAAAATGACCATGTATGAGCAAAAGCAAGCATGTATGATACAAACAAATTTGTGTA 2520
|||||
Db 821 SerAsnGlnLysProCysAspGlnLysGlnTyrGlnLysAspGlnIleGlnThrArgPheValVal 840
QY 2521 GGAACATGAATGGAAAGTGCATCCGGAACACATGCGTAAAGATCAAGAAATGTTATTATG 2580
|||||
Db 841 GlnTyrThrGlnTrpGlnValAspProGlnHisMetArgLysAspHisGlnLysAspValIleMet 860
QY 2581 CCGCTTGAAGAGACGACATATATGATACATCAATTTGGAACATTTACAAAGGATGATCAC 2640
|||||
Db 861 ProTrpArgThrArgHisIleCysThrSerAsnLeuGlnHisLeuGlnThrAspAspHis 880
QY 2641 CCACTTAATGATATATGTGTGATGATTTAGTAAATATCCCTTTTGGCGGATGCTCTT 2700
|||||
Db 881 ProLeuAsnGlnLysAsnIleValIleAspAspLeuValAsnAsnSerPheLeuGlnAspValLeu 900

QY 2701 CTATCAGCAAAATATGAGCAAAACAAAGATTAATACGAATGTATTAAGAAAATTAACSTA 2760
|||||
Db 901 LeuSerAlaLysTyrGlnAlaAsnLysIleIleArgMetTyrLysGlnLysAsnAsnLeu 920
QY 2761 AAGGGCCCCAAGACATCTGACCCAAAACACACAACTATCTGTGAGCTATACST 2820
|||||
Db 921 LysGlnProLysGlnValThrAspProLysHisGlnThrThrIleCysArgAlaIleArg 940
QY 2821 TACAGTTTTCAGATATAGTATATATTCAGAGAAAGACATCTCGGAAAGAAACGGT 2880
|||||
Db 941 TyrSerPheAlaAspIleIleGlnAspIleIleArgGlnArgAspLeuThrGlnAlaGlnGly 960
QY 2881 GACATGCTAAAGCTGCACAGACATTTGAAAATCTGTTTGGTAAATATACATAAGTCACTC 2940
|||||
Db 961 AspMetValLysLeuGlnGlnLysIleGlnLysThrValPheGlnLysHisIleHisLysSerLeu 980
QY 2941 AAAGCAAGCAAAATGATATAATATATGATGATGATGATGATGATGATGATGATGATGATG 3000
|||||
Db 981 LysGlnLysGlnLysAsnAspLysTyrAsnAspAspAlaProLysTyrLeuLysLeuArgGln 1000
QY 3001 AATTGCTGCGAACCTAATAGACCCAAAGTATGCGGAAGCCATGAATGTATATAATAT 3060
|||||
Db 1001 AsnTrpTrpGlnAlaAsnArgAlaLysValTrpGlnAlaIleMetLysCysAspIleLysTyr 1020
QY 3061 TTGAGGATTAATCGGACACCAATCAACAAAGTATGATTTGCGGATATAGTATCAT 3120
|||||
Db 1021 LeuLysAspLysSerGlnHisGlnSerThrGlnSerSerTyrCysGlnTyrSerAspHis 1040
QY 3121 ACACCAATGATGATTAATATATCCACAAAATTAAGATGATGATGATGATGATGATGATGATG 3180
|||||
Db 1041 ThrProLeuAspAspArgTyrIleProGlnLysLeuArgTrpMetThrGlnTrpAlaGlnTrp 1060
QY 3181 TACTGCAAGCTGCAGAAAAGAGATATGATTAAGTGAAGAGAGAGATGATGATGATGATGATG 3240
|||||
Db 1061 TyrCysLysValGlnLysLysGlnTyrAspLysLeuLysGlnLysCysLysGlnLysLys 1080
QY 3241 GATTAAGATTAATGCTCAGGCTGTAGCAAAAGAGTGTACAGGTTGTACAGATGCACA 3300
|||||
Db 1081 AspLysAspAsnGlnGlnGlnCysThrLysGlnSerGlnThrGlnCysThrLysCysThr 1100
QY 3301 GAAGCTTGTAAATTAATTAATGATTAATAGATTAATGATTAATGATTAATGATTAATTA 3360
|||||
Db 1101 GlnAlaCysAsnGlnLysTyrAsnAspIleIleGlnLeuTrpLysGlnGlnTrpAsnIleIle 1120
QY 3361 TCAGATTAATTAACAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3420
|||||
Db 1121 SerAspLysTyrLysGlnLeuHisGlnGlnAlaIleMetSerAlaSerAsnSerGlnIle 1140
QY 3421 GAAGCTTCCAGTACGCCAAAATCATATAGACAGAAATGTAATGTAATTTGTGCGAA 3480
|||||
Db 1141 GlnAlaSerSerThrAlaLysAsnHisIleAspArgAsnValIleGlnPheLeuSerGln 1160
QY 3481 TTAATACCAACAAAATGCTGCAAAAGTAATAAAGTCTATGATTAAGTCTGTC 3540
|||||
Db 1161 LeuTyrGlnGlnAsnGlnLysLysSerAsnLysSerLysThrSerAspGlnSerAlaVal 1180
QY 3541 ATTGCTACTAACACCAAGTATGAAAATGTTGAGAGCATATCTCATGATACAGAAAATTTT 3600
|||||
Db 1181 IleGlnTyrHisAsnThrTyrGlnLysAsnValGlnAlaTyrIleHisAspThrGlnLysPhe 1200
QY 3601 GATGATTTGCACTCACAAAATGAGTTTGTGATGAAAAAAGATGATGATTAAGATTAACGAA 3660
|||||
Db 1201 AspAspCysGlnSerGlnAsnGlnPheCysAspGlnLysSerAspGlnLysAspAsnGln 1220
QY 3661 AAATATGCTTTAGATTAATTAACCAACAGACCAATGATGATGATGATGATGATGATGATGATG 3720
|||||
Db 1221 LysTyrAlaPheArgAspLysProGlnAspHisAspGlnLysLysGlnLysLysSerGln 1240
QY 3721 TCGAAACCGCAAGGCTGATCAATAAAAACGAAAAAAGAAACGGAACGAAAGATTAACGAA 3780
|||||
Db 1241 SerLysProThrArgValGlnIleLysThrLysLysLysLysLysLysLysLysLysLys 1260
QY 3781 TGTAACCACTGATGATTAATTAATTAAGAAAACGATGGAAGAAACAGTGAAGATTTGT 3840

```
Db 1261 CysLysThrValAsnAspIleLeuLysGluAsnAspLysLysGluValGluAspCys 1280
OY 3841 CACCCAAAAAAGATGATATGATATCCGATTTGGCAATCCGGAATATTAATTTAGTG 3300
Db 1281 HisProLysLysAsnSerAsnGlyTyrProAspTrpGlnGlySerGlyAsnIleAsnLeuVal 1300
OY 3901 GAAGACCTCGTGTGTATGATGCCCCAGAAAGCAAAAGTTATGCGTACATTTCTTGCGCA 3960
Db 1301 GluAspProArgValCysMetProProArgGlnLysLeuLysValHisPheLeuAla 1320
OY 3961 AATGATATATGAATTAATAAATTTACAAATCAGATTATTTAAAGAAAGCTTTTCATCAA 4020
Db 1321 AsnAspAsnGlnIleLysLysLeuGlnSerGlnValAsnLeuLysGluAlaPheIleLys 1340
OY 4021 TCTGCAGCAGCAGAAACATCTCTCATGCTATTTATTAATAAGTAAGATGCTGAAGCA 4080
Db 1341 SerAlaAlaIleGlnThrPhePheSerTrpTyrTyrLysSerLysAspGlyGlnGly 1360
OY 4081 AATGAACCTGATTAAGAAATTTAAAGAGCAAAATTCCTCCGCAATTTTGGATCCATG 4140
Db 1361 AsnGluLeuAspLysGlnLeuLysGlnGlyLysIleProProAlaPheLeuArgSerMet 1380
OY 4141 TCTACACATTTGGAGATATATAGATTTTATTTTGAACAGATATATCAAAAGCTCAT 4200
Db 1381 PheTyrThrPheGlyAspTyrArgAspPheLeuPheGlyThrAspIleSerLysGlyHis 1400
OY 4201 GGTGAGGAGAAATTAATAAAGAGCAAAATAGATTTCTTTTCAAAAAGTGGCCAAAAA 4260
Db 1401 GlyGlnGlySerLysLeuLysGlnGlnIleAspSerLeuPheLysAsnGlyAspGlnLys 1420
OY 4261 TCTCTTAATGAAAAACACGCCAAGAAATGCTGACAGAAATAGTCATGATATGGGAA 4320
Db 1421 SerProAsnGlyLysThrArgGlnGlnIleTrpThrGlnHisSerHisGlnIleTrpGln 1440
OY 4321 GGTATGCTATGACTACTATAAATAATGGGCAAAAAAAGATGATTTTCCGAAAATCTAC 4380
Db 1441 AlaMetLeuCysAlaLeuValLysIleGlyAlaLysLysAspAspPheThrGlnLysTrp 1460
OY 4381 GGTACACACACGCTCAATTTAGTGACAAAGCACACACTTTGGAGAAATTTGCCAAAA 4440
Db 1461 GlyTyrAsnAsnValLysPheSerAspLysSerThrThrLeuGlnGluPheAlaLysArg 1480
OY 4441 CCCGATTTTTCAGATGCTTAACCGAATGTCAGACGACTATTTGCTATACAGACAAAA 4500
Db 1481 ProGlnPheLeuArgTrpPheThrGlnTrpTyrAspAspTyrCysTyrThrArgLys 1500
OY 4501 TATTGAAGGATGTGCAGGAAAAATGTATAGTCAAAATGACCAATTGAAAGTGATATCAGA 4560
Db 1501 TyrLeuLysAspValGlnGlnLysCysLysSerAsnAspGlnLeuLysCysAspThrGln 1520
OY 4561 TGTAAATPAAGAAATGCGAGACTAGCTTAATATATCAAAAAAAGAAAGTGAATTCGA 4620
Db 1521 CysAsnLysLysCysGlnAspTyrValLysTyrMetLysLysLysGlnTrpIlePro 1540
OY 4621 CAAGATTAATATTAACAGGATGACGCGACAAAAAAAGTTGATATGACACACATTTGCT 4680
Db 1541 GlnAspLysTyrTyrLysAspIleArgAspLysLysArgPheAspArgGlnHisIleGly 1560
OY 4681 GTAATGTTACAGACTATGAGTGAAGCAATGACAGATTACTTGAACAGAAATTTACT 4740
Db 1561 ValMetValThrAspTyrThrGlyThrAsnAlaThrAspTyrLeuAsnArgLysPheThr 1580
OY 4741 GCTAGTGTGATGAATGAGCTGGAAGTCCCTGTGTGATACAAAGAAATATCAATTTGTTA 4800
Db 1581 AlaSerCysGlyAspLysProGlySerAlaSerValValGlnArgAsnIleGlnLeuLeu 1600
OY 4801 GAAAAACGGCTTACTATGATGCGCAGCAAAATTTGGGTGACAAATTTATTTAGAAAT 4860
Db 1601 GlnLysGlnAlaTyrTyrAspAlaAspLysHisCysGlyCysThrLysPheIleGlnAsn 1620
OY 4861 GACGCAATATATCTAATTAATTTGAGTAAGTAAGTGAAGGTTTGTAAAGAGGCA 4920
|||||
Db 1621 AspAspLysTyrThrAsnIleSerSerLysAspLysCysLysGlyLeuValLysGluAla 1640
OY 4921 AACACAGGTGCTATTAATAGTGGCAAAACAAAGGCTCCTAATTAATTAATTAATTAAGAA 4980
Db 1641 AsnThrGlyAlaIleLysThrPglAsnLysGlyProAsnAsnTyrAsnAsnLeuLysGln 1660
OY 4981 TTGACTGGAAGATGCTGCTTTTCTCTGCTCGACATACGATATATGTTTTCATGATTCAT 5040
Db 1661 LeuThrGluAspValLeuPheProSerArgLeuArgIleCysPheHisAlaLeuAsp 1680
OY 5041 GGCATTTATACAGATCCAGAACTTAAGATGAATAAGTGAATAAGTTCGAAATTAATGGA 5100
Db 1681 GlyAsnTyrThrAspProGlnValLysAspGluAsnGlyLeuArgLysArgLeuMetGln 1700
OY 5101 GTGGCGCCACAGGAAGGTCATTAATTTGGTCAATACATACAAAGAAAAAAGAAAGAG 5160
Db 1701 ValAlaAlaThrGlnGlyTyrAsnLeuGlnGlnTyrTyrLysGlnLysLysGlnLysGln 1720
OY 5161 AAAATPAAGAGTGGATCGGCAAAATATCTATGAGTGGCCGCTGTGATGCTATG 5220
Db 1721 LysIleLysThrSerAspAlaHisLysTyrSerTyrGluValProProCysSerAlaMet 1740
OY 5221 AATATAGTTTATGATTAAGATATTAATTTAGATATGATATTAATTTGGAAGATGAA 5280
Db 1741 LysTyrSerPheTyrAspLeuArgAspIleIleLeuGlyIleAspAsnLeuGlnAspGln 1760
OY 5281 AAACAAAAGCCGAGAAATTTGAAGAAATATTTAAACAAAATGGAACATCAGTGGC 5340
Db 1761 LysGlnLysThrGlnLysAsnLeuLysLysIlePheAsnLysAsnGlyThrSerValGly 1780
OY 5341 AAAGAAAGATAGTACTACAGAAATCCCGTACTACTGCGCGCAAAATTTTTCGAGAC 5400
Db 1781 LysGlySerAspSerThrThrGlnLysProGlySerThrAlaArgLysPhePheThrPsn 1800
OY 5401 GAAAAATGAAGATGTGTGTGGAACGCAATGATATGCGGCTACAAACGTCGTAGGATGAT 5460
Db 1801 GlnAsnLysGlnCysValTrpAsnAlaMetIleCysGlyTyrLysArgGlyValArgAspAsp 1820
OY 5461 GGAATAGTGGAAATAGGCAAGAAAGTGAAGTGAAGTCTTAATAAATAATGCTGTCTACT 5520
Db 1821 GlyAsnSerGlyAsnSerAlaArgSerAspGlnAspLeuLysLysCysGlySerValPro 1840
OY 5521 TCAGATGATGATTTATCCATGATGGGCAAAATCGCGATGAAAGTACTGCGCTATCAGTTCT 5580
Db 1841 SerAspAspAspTyrPrometGlyLysAsnArgAspGlnGlyThrAlaTyrGlnPheLeu 1860
OY 5581 CGATGCTTGGCGAATGGGCTGAAGATTTTGCAAACATTAAGAAAAAGAAATGGAGAA 5640
Db 1861 ArgTrpPheAlaGlnTrpGlyGluAspPheCysLysHisIleLysGlnLysGlnLeuGlnLys 1880
OY 5641 TTGTAAGCGGCTGTAATGATTAATTAATCTGTGCTGTAATGAAGTAAAGAAAGAAATGT 5700
Db 1881 LeuValGlyAlaCysAsnAspTyrThrCysGlyLysAsnGlnLysAspValGlyLysCys 1900
OY 5701 ACAGATGCTGTACACAAATATTAATAAATAATTTATAGTGAAGTGAACCAAGTATGAATA 5760
Db 1901 ThrAspAlaCysThrGlnTyrLysPheIleSerGlnTrpLysProGlnTyrGlnLys 1920
OY 5761 CAATCAAAAAATATGCGTGAAGATTAAGACAAATAATATTCGCGCATCTCTGGCAAAA 5820
Db 1921 GlnIleLysLysTyrGlyGlnAsnLysAspLysIleTyrSerGlnHisProValAlaLys 1940
OY 5821 GATCAGAGAGAGCTCGGAATATTTAGACAAACAAATTAATAAATAATTTGAAATTA 5880
Db 1941 AspAlaGlnLysAspAlaArgGlnTrpLeuAspLysGlnLeuLysLysIleCysGlnAsnLys 1960
OY 5881 AGTGCAGATTGTGAATATTAAGTATGAAGATGTGTCACACAGCATTAATGATGATGT 5940
Db 1961 SerGlnAspCysGlnTyrLysCysMetLysAspValSerThrGlnArgLeuThrAspGly 1980
OY 5941 AATAGTCAAAATATGCGCGCATCTTATGACGATGACCAAAAGAAAGTGAAGAAAGTGT 6000
Db 1981 AsnSerGlnAsnMetProAlaSerLeuAspAspIleProLysGlnValGlnGlyLysCys 2000
```

QY 6001 AATTGTCATAGTCCACAGCTCCACCGTCGTAAGGGAACCCGTCACACGGGCTA 6060
DB 2001 AsnGysGlnValProArgGlyProProArgValArgArgGlnThrProSerProArgVal 2020
QY 6061 TCACGTATCAAAAAGCGAGCGCATGCAAAAAGAACGCAAAACAGCGCGCCACAAAA 6120
DB 2021 SerLeuLeuSerLysAlaThrAlaSerLysGlnAlaLysThrAlaProProThrLys 2040
QY 6121 CAGCCGAAAAAGTGAAGAAATCTAACACAGAAATGCGACACAAACGACCGGAGA 6180
DB 2041 GlnProLysLysValGlnAsnLeuThrThrGlnMetArgAlaGlnThrArgThrArg 2060
QY 6181 GCACGACAAACACACAGCAAAACGACATCAACAGCAACACACAGATCGAGTGGCC 6240
DB 2061 AlaIleGlnGlnThrArgLysArgThrSerThrAlaThrThrGlnSerAspValGly 2080
QY 6241 ACATGGTAAAGCCATCTTCTTGCAATTAACGATAGCAGGGGTGGAATAGCGGTCT 6300
DB 2081 ThrValLysAlaIleLeuSerAsnLysProAspSerArgGlyGlyIleGlnGlyCys 2100
QY 6301 AATCCAAAAAGCATGACACATATCTTAATGGGGTGTATGAGTGAAGTCAAGAA 6360
DB 2101 AsnProLysThrArgLysGlnThrProLysThrPglCysIleValGlyLysSerLysGln 2120
QY 6361 AATGAAAAAGGCATATGTATGCTCTCTCAGAGAAAAAAATTTATGTATTAATATACAA 6420
DB 2121 AsnGlnAsnGlyIleCysMetProProArgArgLysLysLysCysIleAsnAsnIleGln 2140
QY 6421 TATTTAAATATGAACTGAACGAAAAATAGCGACAAATGATATTAAGAGCGCTTATATAA 6480
DB 2141 TyrLeuAsnThrArgLysGlnThrGlnAsnLysArgAspAsnAspIleLysGlnAlaPheIleLys 2160
QY 6481 TGTGCACCAATAGCAAACTCAATTTTGTGGTTAAATATATATATGAAATCCGACAGA 6540
DB 2161 CysAlaIleAlaIleGlnThrGlnPheLeuThrPheLysThrIleGlnAsnProAlaIle 2180
QY 6541 GAAAAATGAAATTCGCAAAATGCAACAATTCAGATGAATTTAAAGAAATATGTATTAACA 6600
DB 2181 GlnAsnGlnLeuGlnAsnGlyThrIleProAspGlnPheLysArgIleMetLysTyrThr 2200
QY 6601 TATGCTATTAATAGATATGTTTGTGGAACGTATTTCTATGATTAATAAAAAATATA 6660
DB 2201 TyrGlnAspLysThrLysAspMetPhePheGlyThrAspIleSerAsnAspLysLysIleIle 2220
QY 6661 ACTGTAAACAATAGTGTACAAACCATCTCATGAAATATATTAAGAAAAACAGATATAA 6720
DB 2221 ThrValIleThrAsnSerValIleThrIleLeuAsnGlnAsnLysLysLysGlnAspLys 2240
QY 6721 AAAAAAGATGAAGAAATTAAGTAAATATTTGGAGAAAAATATAAAAAATTTATTTGGCAA 6780
DB 2241 LysLysAspGlnGlnLeuArgLysIlePheThrPglLysAsnLysLysPheIleThrPglu 2260
QY 6781 GGAATGATATATGATTAATCTATATCATCTACAGACAGAAAAAGAAAAAATTAAGA 6840
DB 2261 GlyMetLleTyrGlyLeuThrThrLysIleuThrAspLysGlnLysGlnLysIleArg 2280
QY 6841 GATATATACAGATACATGACATGACCAAACTGACGCTTCCCTTGAAGAATTGTAAAA 6900
DB 2281 AspAsnThrArgLysThrAsnAspMetThrLysLeuThrProSerLeuGlnGlnPheValLys 2300
QY 6901 AGGCCCAAAATTTTGAATGGTTCACAGAAATGGCGACAGAAATTTTGTATTAAGAGAAC 6960
DB 2301 ArgProGlnPheLeuArgThrPheThrGlnThrAlaGlnGlnPheCysAsnLysArgLys 2320
QY 6961 GAACAGTGTAAATTAAGTGGAGCGCGCTGTAAAGAAATATGAGCTATATGATTAAGAC 7020
DB 2321 GlnGlnLeuLeuLysLeuGlnAlaGlyCysLysGlnLysGlnLysAsnGlySerAsnAsp 2340
QY 7021 GGTAAAGACACAAAGATGTGACAGAGCGCTGTATCATATCAATATTAATTAAGAAATGG 7080
DB 2341 GlyLysThrGlnGlnLysAlaGlnAlaCysValIleThrThrGlnAsnPheIleLysLysThr 2360

QY 7081 AAAACTGAATATGAAAGACAAAGAAAAAGTTCAAAAAAGATATAAGTGCACAAAAGTAT 7140
DB 2361 LysThrGlnThrGlnGlnAlaArgGlnAlaArgGlnLysPheLysLysAspLysPglLysLysTyr 2380
QY 7141 AAGCATATATCTTCTACTGAAAGACATAGAGAACCAACATGCTGCTCATGAATATTA 7200
DB 2381 LysAspLysProSerThrThrGlnAlaArgAspIleGlnLysAlaThrCysAlaIleGlnThrLeu 2400
QY 7201 AACATGAATTAAGAAATTAATGAGCAATTAAGCATTTGTTGTATGACAAAAACCTTCT 7260
DB 2401 AsnMetLysLeuLysGlnLeuCysGlyAsnLysAspCysSerCysMetGlnLysProSer 2420
QY 7261 TCACAACTACCAAAAACACACACAAATCAACATCATCCAGTGTATGATATGCCAGAA 7320
DB 2421 SerGlnLeuProLysThrThrGlnGlnSerGlnSerSerAspAlaAsnAspMetProGln 2440
QY 7321 TCCGTGATATATGTTCTCGAAGAAATTAACAAAGTGAGTGCTCGAATTTCAAAAAAG 7380
DB 2441 SerLeuAspLysValProGlnGlnPheAsnLysCysGlnCysProGlnLeuSerLysLys 2460
QY 7381 GGAATCATATGATTCATACAAAAAAATTAACCTTAACCTTAATGATGTGTAGAG 7440
DB 2461 GlySerMetIleHisThrLysLysIleThrGlnProLysIleProMetLysCysValGln 2480
QY 7441 AAACGACCATATTTATATCTAAAGAACGCAAGAAATATATGATATTAATACCTGAAGAA 7500
DB 2481 LysAlaIleAlaTyrThrLeuSerLysGlnAlaGlnAsnMetAspIleThrLeuLysGln 2500
QY 7501 AAATTTACTCTATGAGTGTACAAAGAAAGAAAGTAAATATGTGACTAATATAT 7560
DB 2501 LysPheIleProIleGlnLysThrLysGlnLysLysSerLysAsnSerThrPheAsnAsn 2520
QY 7561 AATCCTTGCAATCTTAAGAAACCTTATGACACCTGATTAATATATAGGAAAGAAACCTT 7620
DB 2521 AsnProCysAspProLysLysProThrAlaProAspLysThrIleGlnArgAlaAsnPro 2540
QY 7621 TGTGAAAAATGACAGAAAAATCGTTTAAAGGTAGATATTAAGAAATGTTACAAAAT 7680
DB 2541 CysGlnAsnArgGlnGlnAsnArgPheLysValAspThrGlnThrLysCysTyrLysAsn 2560
QY 7681 TCAAGTGTATCATCGAGAAAAAAAGATGTGTACTCTCAAGAAAGAAACATATGTGC 7740
DB 2561 SerLysPheThrGlnGlnLysLysArgValCysValProProArgArgGlnHisMetCys 2580
QY 7741 TTAAGAAATTTAGATGAATTAATAATTAAGAAAGCTTAAGGATAGTAAATTTCTCTAAA 7800
DB 2581 LeuArgAsnLeuAspGlnIleLysIleGlnAlaArgLeuLysAspSerAsnThrLeuLeuLys 2600
QY 7801 ATGGTCTGTCGACACTGACAGAAATGAAGAAATGACATTAATAAAAAATTCACACTCAGAG 7860
DB 2601 MetValAlaArgArgThrAlaAlaArgAsnGlnGlyIleAspIleIleLysAsnPheAsnSerGln 2620
QY 7861 AACGGGTGCGCAATGAATTCACAAATATGATGATGATTAATTAATTTGCTGATCTGGGT 7920
DB 2621 AsnGlyCysAlaMetAsnProIleCysAspThrMetLysThrSerPheAlaAspLeuGly 2640
QY 7921 GACATAGTTAGAGAACAGATATGTTACGAATTTGGTGGTACTTACCCTCCGTAAGAATA 7980
DB 2641 AspIleValAlaArgGlnThrAspMetLeuArgIleGlyLysThrLeuProProValGlnIle 2660
QY 7981 AAATTAATTAAGCTTTTGAATATATATATGAAGAAATGAAGAAATTAAGAGTAGA 8040
DB 2661 LysLeuThrLysValPheGlnThrIleTyrGlyLysThrArgAsnLysAsnLysGlyArg 2680
QY 8041 AATTAATTAACAAGATGTAACAAGCTTCTGCTGCTGGTGGGATGCTATTAAGAAAAAGAT 8100
DB 2681 AsnLysThrAsnAspValGlnThrPheArgSerLalaThrPheAspAlaAsnArgLysAsp 2700
QY 8101 ATTTGGAAGCAATGACGTGCAAGACACAGAGATGCAAAAATTTTATAGAAAAAGAGA 8160
DB 2701 IleThrPysAlaMetThrCysLysAlaProGlnAspAlaLysLeuPheArgLysGlyArg 2720
QY 8161 ATGATGCAATTTGAACGCATTAACATTAATACAAAGATTAAGTGTGACATTAAGAGCATCA 8220

Db 2721 MetaspGlyPheGluArgIleThrLeuIleGlnAspLysCysGlnHisLysAspAspPro 2740
QY 8221 CCTGTGATGATTTATATACCTCAACGGTTTGATGATGACTGAATGGTCGATATATAT 8280
Db 2741 ProValnAspAspTyrIleProGlnArgPheArgTyrPheThrGlnTyrSerGlnTyrTyr 2760
QY 8281 TGTAAACCACTGATGAGAAATTTGAAAAATTTAAAAATCATGTGATCATCTGTAACA 8340
Db 2761 CysLysAlaLeuMetGlnGlnLeuGlnPheLysLysSerCysAspHisCysLysThr 2780
QY 8341 TGTGACGATGCAAGATGATTTATGAGAAAAATTAAGTGTGAACAGTGAACAGATGT 8400
Db 2781 SerAspTyrCysLysAsnAspTyrAspGlnAsnLysCysGlnGlnCysLysThrArgCys 2800
QY 8401 CAAGAATATAAAAATTTGTTCTTAATGGAATCTGATTCGATATACAAATCAATAAA 8460
Db 2801 GlnGlnTyrLysAsnPheValLeuLysTyrLysSerLeuPheAspIleGlnSerAsnLys 2820
QY 8461 TACAAAGAAATTTGATGAAACAACAAATATATACAAAAATCTACTATGATCATGTGCA 8520
Db 2821 TyrLysGlnLeuTyrGlnGlnProIleTyrThrLysIleSerThrTyrAspHisValGln 2840
QY 8521 AATTTGTACAAAAGTTGAAAACCTTTAAAGAGATGTTCTGTGAGACTTTCTGGA 8580
Db 2841 AsnPheValGlnLysLeuLysThrPheLysSerGlnCysSerValGlnSerPheSerGln 2860
QY 8581 TATCTTCAGAAAACAAGTAAGTGTTCGAATTAATTAATTAATGAATGAATGTTCTTC 8640
Db 2861 TyrLeuHisGlnThrSerLysCysLeuAsnTyrLysPheAsnGlnAsnAspGlySerSer 2880
QY 8641 AATATACGAACATATGCTTCGAGAAACCCCAAAAGTTATTAAGAAGCTTCAGTTGT 8700
Db 2881 AsnIleArgThrTyrAlaPheGlnGlnThrProLysSerTyrLysGlnLacCysSerCys 2900
QY 8701 ACACACTCTCTTAAGATCCATGATTAATTTGCTACCGATCAAAACAAGATGATG 8760
Db 2901 ThrLeuProSerLysAsnProLeuAspAsnCysProThrAspGlnAsnLysAspGlyCys 2920
QY 8761 AAGCAATTCACAACTTTTACCTTCCTCTCGAAGATGATTAATTAATTAATCTTGATTA 8820
Db 2921 LysGlnLeuGlnThrPheThrPheCysSerLysAsnAspTyrLysAsnAsnLeuAspAsn 2940
QY 8821 TGGACGATACCTGTTCTTATATAGTCAGATGATTAACAAGGCTATGATTCCTCA 8880
Db 2941 TrpAsnAlaTyrLeuValIleuAsnSerSerAspAspAsnGlyValLeuIleProPro 2960
QY 8881 AGAAGAAGACATTTATGTAACAAGCTATCACTGCATATATTAATTAAGAAGGTGATAA 8940
Db 2961 ArgArgArgHisLeuCysThrArgProIleThrAlaTyrAsnTyrArgLysGlnAspLys 2980
QY 8941 GAAATTTTAAAAAAACTTCTTACTTCTGCTTCAGTCAGCAAGACAATTTGATAGTCA 9000
Db 2981 GlnIleLeuLysLysLysLeuLeuThrSerAlaPheSerGlnGlyGlnLeuLeuGln 3000
QY 9001 AAATATAAATCGAAGAAAGTGTGCTTGAAGCAATGAATATATATGATATGATAT 9060
Db 3001 LysTyrLysSerGlnGlnLysLeuCysPheGlnAlaMetLysTyrSerTyrAlaAspTyr 3020
QY 9061 TCCGATATATTAAGAACTGATATGATGACACTTCATTAATCTGAAAAATTAATAAA 9120
Db 3021 SerAspIleIleLysGlnThrAspPheMetAspThrSerLeuSerGlnLysIleLysLys 3040
QY 9121 AATTTGAAACATCAATGAAGAACCCGAAATCGTAAACATGGTGGGAAATTAATGA 9180
Db 3041 IlePheGlnThrSerAsnGlnAlaThrGlnAsnArgLysThrTyrPheGlnAsnAsnArg 3060
QY 9181 CGTCAGATATGGCAGCTATGTTATGAGATTAATAATGCTACTCTCAAAAGATACATTA 9240
Db 3061 ArgGlnIleTyrPheHisAlaMetLeuCysGlnTyrLysIleAlaThrSerLysValThrLeu 3080
QY 9241 CATGAAGATGGTGTCAATTAACCAAGATGAAGAACTAATCACTTCTCTGTTGTTA 9300
|||||

Db 3081 AspGlnGlyTyrPheCysGlnLeuProLysAspGlnGlnThrAsnGlnPheLeuArgTyrPhe 3100
QY 9301 ATTTGAATGGCAAGACAGCATGTAAGAAAAACATGTAAGTTCATTTAAAAACA 9360
Db 3101 IleGlnThrPalaLysGlnAlaCysLysGlnLysLysHisValSerAspSerLeuLysThr 3120
QY 9361 AAATGTCCTGTCGAACGAGATTAATTTGAAGGCTCGAGATTTATTAAGACACCTGGA 9420
Db 3121 LysCysProLysSerAsnLysAsnAspPheGlnAlaSerGlnLeuLeuArgGlnProGly 3140
QY 9421 TGTCAAGATGATTAATTAAGAAATATATTAATTAATTAATTAATTAATTAATTAATTA 9480
Db 3141 CysGlnAsnAspIleArgLysTyrIleSerLeuAsnIleLeuIleLysAsnThrMetGln 3160
QY 9481 AATCTAAATTAATAATTAAGCAATTAATAAGATCAATCTTCAGGTATATATACAAATAA 9540
Db 3161 AsnLeuAsnIleLysTyrLysGlnLeuLysAspGlnSerSerGlyAsnIleAspAsnLys 3180
QY 9541 CCATCGAAGAAAAATGTCAGTCATATTAATAATCAAAAAGATTCGATCGCTTGGAG 9600
Db 3181 ProSerGlnGlnAsnValGlnInserTyrIleLysSerLysAspSerGlnCysAlaLeuGln 3200
QY 9601 TTAATGATATTAATGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9660
Db 3201 LeuAsnAspIleAsnGlnIleValThrGlyThrLysAsnAsnGlnLysAsnGlnLys 3220
QY 9661 GAAGTACTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9720
Db 3221 GlnValLeuLysLysLeuTyrProGlyLeuTyrPheValGlnAspGlnThrHisLysAsn 3240
QY 9721 CATGTACTAGATGAAATTAATAAGAAAGAAAGCAAGATTCGCTCAAAAGCACTGAT 9780
Db 3241 HisValLeuAspGlyAsnIleLysGlnGlnGlnGlnThrValArgProLysAlaLeuTyr 3260
QY 9781 TTTCTTAACCCCATGATGATTTCTTATGATCAAGCACTTATTTCTGCAACATGAGTA 9840
Db 3261 PhePheThrProHisValAspSerPheTyrGlnAlaProLeuPheSerThrHisArgVal 3280
QY 9841 GCACATATGATCCTTAATAATGATATTAATTAATTAATTAATTAATTAATTAATTAATTA 9900
Db 3281 AlaGlnTyrAspProLysAsnAspIleLeuLysSerSerIleSerValIleAlaSer 3300
QY 9901 GCGTTAGTTGATAGCGCTTATATTCATGAGAAAAAAATCAATCGTCTGGACTTG 9960
Db 3301 AlaLeuGlnLeuIleAlaLeuHisPheMetLysLysLysPheLysSerSerValAspLeu 3320
QY 9961 TTGCTATACGAATATCCCGAAGAGATGAAAGCTTACGTTGAGATCCAAAAAT 10020
Db 3321 LeuArgIleLeuAsnIleProGlnGlnGlnLysGlnLysMetProThrLeuGlnSerLysAsn 3340
QY 10021 AGTACATACCATATAGAAGTGTCCATATAGAAGCAAAACATTAATATATATGAGAAAGA 10080
Db 3341 ArgTyrIleProTyrArgSerGlyProTyrLysGlnTyrThrTyrIleTyrMetGlnGly 3360
QY 10081 GATCTACTGAGATGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10140
Db 3361 AspThrSerGlyAspGlnLysAspLysTyrMetThrPheLeuSerSerAspIleThrSer 3380
QY 10141 TCCGAAGTGAAGTGAAGATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10200
Db 3381 SerGlnSerGlnTyrGlnGlnLeuAsnAspIleAsnAspIleTyrValProGlySerProLys 3400
QY 10201 TATAAAACATTAAGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10260
Db 3401 TyrLysThrLeuIleGlnValIleValLeuGlnProSerLysArgAspIleProSerAspAsp 3420
QY 10261 ACACCAAGTATGATTAACCAACGATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 10320
Db 3421 ThrProSerAsnAspThrProArgThrAsnArgPheIleAspAspGlnTyrAsnGlnLeu 3440
QY 10321 AAACATATTTTGTATGCAATATTTACCAAAATGACAAACCAATTAATTAATTAATTAATTA 10380
Db 3441 LysHisAspPheValSerGlnTyrLeuProAsnThrGlnProAsnAsnTyrLysSer 3460
|||||

QY 10381 GCAGATATTCACATGATACAGACCTAATACCTTATATTCGATATATCCTGAAGAAA 10440
 DB 3461 Alaspielerproketastrnhtluproasthlleutyserspansprogluglulys 3480
 QY 10441 CCTTTATATATATTCATTCATGATAGGATTTATATACCTGGAGAAATATCTATAT 10500
 DB 3481 Prohellelleleserillestisaspargaspseutyrthrlglyserglyserlyasn 3500
 QY 10501 ATTAATATGCTACTAT 10560
 DB 3501 Ileasmetserthrasnthrasnasnaspielerprometnasnalaargasnaspsertyr 3520
 QY 10561 AGAGTAT 10620
 DB 3521 Argglylehspleuleisnaspsertleuvalleuleusnleuleuiletyrmetmet 3540
 QY 10621 AAGTAT 10626
 DB 3541 Lystyr 3542
 ID AAM22475 standard; Protein; 3060 AA.
 AC AAM22475;
 DT 12-SEP-1997 (first entry)
 DE Plasmodium var-7.
 KW DBL gene family; SABP, stalloic acid binding protein; vaccine; therapy;
 KM Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KM DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KM Plasmodium.
 OS Plasmodium vivax.
 OS Plasmodium falciparum.
 PN MO9640766-A2.
 PD 19-DEC-1996.
 PF 07-JUN-1996; 96WO-US09508.
 PR 07-JUN-1995; 95US-0487826.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 CH Chitnis C, Miller LH, Peterson DS, Slim KL, Su X;
 WE Wellens TE;
 DR WPI: 1997-052231/05.
 DR N-PSDB; AAT72882.
 PT New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and stalloic acid
 PS binding proteins
 PS Claim 8; Page 61-67; 96pp: English.
 CC This sequence represents var-7 of Plasmodium. Var-7 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and stalloic acid binding protein
 CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 CC var family of genes modulate cytoadherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and
 CC SABP mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the
 CC invention. The compositions are for the treatment and prevention of

CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
 CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
 CC family of genes having homology with conserved regions of DABP and SABP.
 CC The compositions are used for the treatment and prevention of malaria.
 CC They are also used in the preparation of vaccines for inducing a
 CC protective immune response in a mammal to Plasmodium merozoites
 CC (especially Plasmodium falciparum or Plasmodium vivax).
 CC XX
 SQ Sequence 3060 AA;
 Alignment Scores:
 Pred. No.: 5,39e-238 Length: 3060
 Score: 3501.00 Matches: 1057
 Percent Similarity: 38.65% Conservative: 470
 Best Local Similarity: 26.75% Mismatches: 934
 Query Match: 18,008 Indels: 1490
 DB: 18 Gaps: 147
 US-10-087-013-1 (1-10628) x AAM22475 (1-3060)
 QY 25 ATTATATAAATGGGATGCGATCATCATCTATAGAGGAGATGCTAAGCCCTATTATA 84
 DB 4 LeuAlalysketcllyProlysgluAlalalgllyAspAspilleglusap----- 20
 QY 85 AAGAAAGTCACAAAGTCACAAAGTGTTCGAAAGCTATATCCAAATATATAGACAT 144
 DB 21 -----gluseralalysShlsmetPheAspArgllyleglylsaspValTyrAsp 36
 QY 145 CCATCAAAA---TATGCAAAAGAACATGTGATTCGTGAAAGGATTTGAGCAAGCA 201
 DB 37 LysvallysgluAlalalysgluAlargglyLysglyleuendInglyrleuSerGluAla 56
 QY 202 GAATTCGTGCTGCTCTTACGCCAGTAAATAGATATATATATATATATATATATATAT 261
 DB 57 LysPheglulysasnleuSerAspProglInthProglusap-----Pro 71
 QY 262 TGTAAATTTAGATCATAGAACATCTAATTTAGCGTATGATGATGATGATGATGATGAT 321
 DB 72 CysAspPheunsphtllyLysTyrHlsthAsnVal-----ThrThrAsnValIleasn 88
 QY 322 CCTTGCCATGTGAGAACAAACCGATTTGATGAAAGATGAAAGATGATGATGATGATGAT 378
 DB 89 ProCysAlasparlySerAspValArgPheSerAspLulyllyllyllylncysThrHls 108
 QY 379 AATAAATACGTAATATATAAAGAAAAATGATGATGATGATGATGATGATGATGATGATGAT 438
 DB 109 AsnArgglylehspsersergInglyAspAspAspAspAspAspAspAspAspAspAsp 128
 QY 439 CGACATATGTGTATATAAAGCTTGAAGCTCTAATGATATATAATATATATATATATAT 498
 DB 129 LeuHlvalCysAspInAsnleuGluInllegluproIlelylthethrAsnThrHls 148
 QY 499 GATTATATGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 558
 DB 149 AsnleuValAspValCysmetAlaAlaLysPheInllyInserlIethrInAsp 168
 QY 559 CATCA---CATAAGGACT-----TCAGACCTTACTGCTGCTGCA 600
 DB 169 TyrProlytyrGlnAlaThrlyrGlyAspserProserInllycysthrMetleuAla 188
 QY 601 CGAAGTTTCCAGATATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 DB 189 ArgSerPheAlaspielleglyAspIleValAlargglyArgaspseutyrleuGlyAsnPro 208
 QY 661 CATGAC-----AAGTAGAAGACGGGTCTCGAAGGTTTCAGAAATA 705
 DB 209 GlnGluIlelysglnArggInglyInleuGlnAsnAsnleuysThrIlePheglyLysle 228
 QY 706 CATGAT-----GGAATGGAAGATGAGTAAATATGATATATATATATATATATATATAT 756
 DB 229 TyrgluLysleuasnnglyAlaGluAlaArgTylGlyAsnsp-----Proglu----- 244
 QY 757 GGAATTTATATATAATTAAGACAGCATGCTGGAATGCTGAATGCAATAATTAATGAGGAA 816

Dd	596	---	tyllsgclglsuaspval	val	lysval	lglnhsasrglsuaspasrgl	ugl	uasprty	614
Qy	1741	GATATACCTAAG	-----	ATAATA					175
Dd	615	glusnrvllayasnmla	gl	ugl	ueu	Cyslleleu	lysasn	gln	lys
Qy	1759								
Dd	635	glugl	ugl	ysan	thr	ser	gl	uysg	lup
Qy	1774	TTTGAATATGAGTGCATATTAATTAAGGATACATTAATGAGTAATGACAACTTAA							183
Dd	655	pheuytyl	trp	ral	ala	his	met	leu	ysas
Qy	1834	ACTTGATA	---	ATAATACAA	CCAGCATGT	---	ATTGATGAATGTACAGAAATTC		188
Dd	675	argcys	leu	gln	asn	gln	sn	l	ysn
Qy	1888	TTTATGTTTTCACAGATGGTTTAAACAAAGAAAGAAAGAAATGAAATGAAATGAAATG							194
Dd	695	glucys	phel	ys	arg	thr	leu	thr	gln
Qy	1948	TTTCAAAAAA	AAAAAGAAATACAG	-----					197
Dd	715	Phe	---	lyst	thr	gln	asn	lle	ys
Qy	1972	-----	CAATCATTTATAGTAATTTATATATCTTTTGAAGTTATTT						201
Dd	734	Pro	ph	as	ph	is	as	pr	ly
Qy	2017	-----	TTTAACTTATGATTAACCTGTGCACAAAGATGACAAAGAAATG						205
Dd	754	ser	gl	u	as	ph	al	as	er
Qy	2059	AAAGAACTTATGAAATATTTAAAGAAAAA	AAAAATGATTTTCAATTGGAAAAATAT						211
Dd	772	---	glu	gln	leu	h	is	leu	arg
Qy	2119	AGGAC	-----	TATTTAGAAATGCATTAACACTCTTGTA					215
Dd	788	gln	gln	u	al	as	er	val	gln
Qy	2155	GATCACTTAAAGAAACTGCCAGATATGT							218
Dd	808	asn	ty	gln	u	as	ph	al	as
Qy	2185	---	AAAGACATTAATACAAAGGAAGCATGTGAAACATCCATATATGACAAACAAACCG						224
Dd	828	lys	gln	u	ys	gln	as	ph	al
Qy	2242	TGTGTAAACCTCGTAGAGCAGCA	---	CCCACTAAATATTAAGAAATACACAA					229
Dd	848	Cys	ser	gln	u	as	ph	al	as
Qy	2299	TACTTAAACGATGCAATGAGAAAGCAACGAATCGTGCTTCATTAATTTGAAGCA							235
Dd	868	met	his	his	lys	ala	lys	thr	gln
Qy	2359	AAGGACACGAAGGTATATTAACGTTGGGGGTAGAGAAAGCACTTCAGACAAATTTA							241
Dd	888	asp	leu	ser	leu	al	gln	phel	ys
Qy	2419	TGTAGATTAATGATTAACATCTTAATTCGTAATCTTGTTTTCAAATGACAACTGTAT							247
Dd	908	Cys	lys	leu	asn	gln	u	as	ph
Qy	2479	GGCAAAAGCAGAGTGATGATTAACAACAAGATTTTCGTAGCAACGAAATGCGAATG							253
Dd	928	gln	lys	---	asp	leu	as	ph	al
Qy	2539	GATCCGAACACATGCGTAAAGATTCACAGAAAGATGTTATATGCTCCCTAGAAAGCAGAT							259
Dd	947	leu	gln	u	al	u	ys	gln	h


```
Db 1644 AspProGluTyrLysGlyTyrGluTyrLysAsp----- 1654
QY 4723 TTGAACAGCAAAATTACTAGTAGTTGGTGATTAAGCCTGGAAGTCCCTCTGTGTACAA 4782
Db 1655 -----GlyValGlnPro-----IleGln 1660
QY 4783 AGAAATATACAAATTGTTAGAAAAACAGGCTTACTATGATGCCGCAACAATGTGGGTGC 4842
Db 1661 GlySngluTyrLeuGlnLys-----CysAspSnsAsnLys---CysSerCys 1676
QY 4843 ACAAATTTATTTAGAAATGACGACAAATATCTAACATTTTCAGCTAAAGATAGTGCAA 4902
Db 1676 ----- 1676
QY 4903 GGATTAGTAAAGAGCAACACAGTGTCTATTAAAGTGCAAAACAAGGTCCCTAATTAC 4962
Db 1676 ----- 1676
4963 TACAATTACTTGAAAGAAATTGACTGAAGATGTGCTTTCTCTGTCGACTACGTATA 5022
1676 ----- 1676
QY 5023 TGTTCATCATGATGATGCAATTATACAGATCCAGAAATTAAGATGAATGGGTG 5082
Db 1677 -----MetAspGlyAsnValLeuSerValSerProLysGluLys----- 1689
QY 5083 CGAAAAAGATTGATGGAAGTGGCGGCAACGGAAGGTACAAATTTGGGTCACTACTACAA 5142
Db 1690 -----ProPheGlyLysTyr----- 1694
QY 5143 GAAAAAAGAAAAAGAAAAATTAACGTCGGATGCCGACAAATTTCTTATGAGCTC 5202
Db 1695 -----AlaHisLysTyrPro----- 1699
QY 5203 CCGCCTTGATAGTCTATGAATATAGTTTATGATTAAAGATATTAATTCTAGTATT 5262
Db 1699 ----- 1699
QY 5263 GATAAATTTGSAAGATGAAAAACAAGAACCCGAGAAATTTTGAGAAATATTTTAACAA 5322
Db 1699 ----- 1699
QY 5323 AATGGAACATCACTGGCAAGAGAGTATGACTACAGAAATCCGGTACTGCG 5382
Db 1699 ----- 1699
QY 5383 CGAAATTTTCTGAGAGAAATAAGAAATGTGTGTGGAAGCAATGATATCGGGTAC 5442
1699 ----- 1699
QY 5443 AAACGTTGAGGAGATGATGAAATAGTGAATAGTGCAGAAGATGATGAATCTAATA 5502
Db 1699 ----- 1699
QY 5503 AAATGTGTTCTGTAACCTTCAGATGATATATCTATGAGGAAAAATCGCATGAAGT 5562
Db 1699 ----- 1699
QY 5563 ACTGGTATCACTTCTTCGATGCTTTGCCGATGGGTGAAGATTTTTCAAACATAA 5622
Db 1699 ----- 1699
QY 5623 GAAAGGAATTGAGAAATTTGTTAGCGCGCTGATATTAATCTTGCTGATATGA 5682
Db 1699 ----- 1699
QY 5683 GATTAAGAAAGAAATGTACAGATGCGGTACACAAATATTAATAATTTATTAGTAGTGG 5742
Db 1700 -----GluLysCys----- 1702
QY 5743 AAACACAGTATGAAAAACAATCAAAAATATGTGTGGAATTAAGACAAAATATATTCC 5802

Db 1702 ----- 1702
QY 5803 GACATCTCTGTGGCAAAAGATGACAGAGACGCTCGCAATATTTAGACAACAATTAATA 5862
Db 1702 ----- 1702
QY 5863 AAATTTTGAAAAATTAAGATGAGATTTGTGAATATTAAGTATGAAGATGTGCCACA 5922
Db 1703 -----AspCys----- 1704
QY 5923 CACCGATTAACTGATAGTAAATAGTCAAAATATGCCCGATCATTTAGAGATGAACAAA 5982
Db 1704 ----- 1704
QY 5983 GAAGTTGAAGAAAGTATTAATTGTCAAGTCCGACGAGGTCCACACGTGTACGAAGGAA 6042
Db 1705 -----TyrGlnGlyLys-----HisValProSerIleProPro----- 1715
QY 6043 ACACGCTCACACGGGTATCACTGATATCAAAAAGCAGCGCATCCAAAAAGACGAAA 6102
Db 1715 ----- 1715
QY 6103 ACAGCGCCGCTACAAAACAGCCGAAAAAGTGAAAAATCTAACACAGAAATGCGACA 6162
Db 1716 -----ProProProProValGlnProGln----- 1723
QY 6163 CAAACACGACCCGACGACGACGACACACAAACAGAAAAACAACATCAACAGCAACA 6222
Db 1724 -----ProGluAlaProThr----- 1728
QY 6223 ACAGATCTGACGTGGGACATGTTAAGGCCATCTTGCAATTAACAGATAGCAGG 6282
Db 1729 ValThrValAspValCysSerIleValLysThrLeuPheLys-----AspThrAsn 1745
QY 6283 GGTGAATAGAGCGTTGTAATCCAAAAACGTATGACAAAT-----CTAAATGGCGT 6336
Db 1746 AsnPheSerAspAlaCysGlyLeuLys-----TyrGlyLysThrAlaProSerSerThrLys 1764
QY 6337 TGTATT-----GTAAGTAAGTCTAAAGAAAT 6383
Db 1765 CysIleProSerAspThrLysSerClyAlaGlyAlaThrThrGlyLysSerGlySerAsp 1784
QY 6384 GAAATGCAATATGATCCCTCTAGAGAAAAAATATGATTAATTAATATCAATAT 6423
Db 1785 SerGlySerIleCysIleProProArgArgArgLysLeuGln----- 1803
QY 6424 TTAATTAATGAAC-----GAAATACGCTGAC 6453
Db 1804 -----GluThrAlaThrAlaLeuProGlnGlyGluAlaAlaProSerThrIleSerArgAla 1822
QY 6454 AATGATATTAAGAGCGCTTTATTAATGATGTCAGCATAGAACTCAATTTTGTGTTA 6513
Db 1823 AspAspLeuArgAsnAlaPheIleGlnSerAlaAlaIleGluThrPhePheLeuThrPasp 1842
QY 6514 AAATATATTAATTGA-----AATCTGCACAGAAAAATGAA----- 6549
Db 1843 ArgTyrLysGluGluLysLysProGlnGlyAspLysSerGlnAlaLeuSerGlnLeu 1862
QY 6550 -----TTGCAAAATGGAACA 6564
Db 1863 ThrSerThrTyrSerAspAspGluGluAspProProAspLysLeuGlnAsnGlyLys 1882
QY 6565 ATTCCAGATCAATTTAAAGAAATATGATTAATGATGATGATTAAGATGATGTT 6624
Db 1883 IleProProAspPheLeuArgLeuMetPheTyrThrLeuLysAspLysPyrArgAspIleLeu 1902
QY 6625 TTGGAACGTATTTCTATGATTAATAAATTAATATTAATCTATACAAATAGCTGAACACC 6684
Db 1903 ValHisGlyLysThrSerAspSerGlyAsnThrAsnGlySerAsnSnsAsnIle 1922
QY 6685 ATTCTCAATGAATAATATTAAGAAAAACAGATAAAAAAGATAGAGAAATTAAGTACATA 6744
Db 1923 ValLeuGluAlaSerGlyAsnLysGluAspMetClnLysIleGlnLysIleGln 1942
```

[illegible]

OY	7501	AAATTATTA	CTATTGAGTCTACAAAAGAAAAAGAAAGTAAATAATTCGTGGCAATAAT	7560	
Dd	2231	-----	 ::: :: LysAspSglucyAspAsnSer	2238	
OY	7561	AATCGTTGCATCCTA	TAGAACAACCTTATGCACCTGATAAATATATAGGAAGAAACCCCT	7620	
Dd	2239	-----	::: LysGlyThrAsp	2242	
OY	7621	TGTGA	AATATACAGAAAAATCGTTTTAAAGTAGATTATGAATGGAAAATGTTACAAAAT	7680	
Dd	2243	CysArgAsn	-----	Lysasn 	2247
OY	7661	TCAAAGTTC	TACGAGAAAAAAGAGTGTGTCTCCACAGAAGAACATATGTGC	7740	
Dd	2248	Ser	-----	2248	
OY	7741	TTAAG	AATTTAGATGAATTTAAATTTGAAGACTTAAGATAGTAATTAATCTCTAA	7800	
Dd	2249	-----	::: :: IleAspAlaThrAspIleGluAsnGlyValAspSerThrValLeuLueet	2255	
OY	7801	ATGCTGCTCGA	CTGACGACCAAAATGGAAGMATAGCATATATAAAACTTCAACTCA	7857	
Dd	2266	ArgValSerAlaAspSerLysSer	-----	GlyPheAsnGlyasp 	2278
OY	7858	-----	GAGAACGGGTCCGCAATGAATCCAAATGTGATACATACAAATATAGTTTGCT	7911	
Dd	2279	GlyLeuGlnAsnAlaCys	-----	2284	
OY	7912	GATCGG	GTACATGTTAAGGAAACGATATGTACAAATTTGGTGGTACTTACCCTCC	7971	
Dd	2285	-----	 ArgGlyAlaGly	2288	
OY	7972	GTGAA	ATAAAAATTTATTAAGGTTTTTGAATACATATATGAAAAATGAGAAATAAAAAT	8031	
Dd	2289	-----	::: :: IlePheGluGlyIle	2293	
OY	8032	AAAGT	TACAAATTAATACAAACGATGTACAAACGTTTCGTTGCTTGGTGGAGTCTAAT	8091	
Dd	2293	-----	2293		
OY	8092	AGAAA	AGATTTATTCGAAGCA-----ATGACGTGCAAAAGCACCAACAA	8133	
Dd	2294	ArgLysAspGluTrpLysCysArgAsnValCysGlyTyrrValValCysLys	--ProLiu ::: :: 	2312	
OY	8134	GATCG	AAAACTTTTACAAAAGAGACAGATGATGATTTGAACGCATTAACATTATACAA	8193	
Dd	2313	AsnValAsnGlyGluAlaLysGlyLys	-----	2321	
OY	8194	GATTA	GTGTGACATTAAGACGATGCCACCTGTTGATGATTATATACCTCAACGGTTTGA	8253	
Dd	2322	-----	::: :: HisIleIleGlnIleArgAla	2328	
OY	8254	TGATG	ACTGAATGCTCGTAATTTATATTGAAAGCACTGATGGAAAGATTTGAAAAATTT	8313	
Dd	2329	LeuValLysArgTrpValGluTyrrPhe	-----PheGluAspTryrAsnLysIle	2344	
OY	8314	AAAAAT	CATGTGATCAGTCGTAAACAACCTCGACAGATGCAGAAAGATATATATGATGA	8374	
Dd	2345	LysHis	-----LysIleSerHisArgIleLysAsn-----Gly	2355	
OY	8374	AAGTGT	GCACAGTAAACGAGAGTGTCAAGATATATAAATTTTGTCTTAATATGAAA	8433	
Dd	2356	GluIleSerProCys	-----IleLysAsnCysValGluLysTrp---	2368	
OY	8434	TCTCAT	TTCGATATACAAATTAATACAAAGATTTGTATGACACACCAATATATAC	8493	
Dd	2369	-----ValAspGlnLysArgLysGluTrpLysGluIleThrGluArg	-----	2382	
OY	8494	AAATCT	CTACTATATGATCATGTTCAAAAATTTGTACAAAAGTGTAAAACTTTAAAGT	8553	
Dd	2383	-----	PheLysAsp	2385	
OY	8554	GAA	TGTTCTGTGAGAGCTTTTCTGATATATCTTCATGAAGAACAGTAAAGTGTTCGATTAAT	8613	


```
Db 2386 Gln-----Tyr 2387
QY 8614 AAATTAAGAAATGATGTTCTTCCAATATAGAACATATGCTTCGAAAGAACCA 8673
Db 2388 Lys---AsnAspAsnSerAspAspAsnValAlaIleSerPheLeuGlnIlePro 2406
QY 8674 AAA-----AGTATTAAGAAGCT 8691
Db 2407 GlnIlePheAspAlaAsnAlaLysAsnLysValIleLysLeuSerLysPheGlyAsnSer 2426
QY 8692 TGCAGTTGTACACACTCTTCAAAATCCATGATTAATGCTCTACCGATCAAAACAA 8751
Db 2427 CysGlyCysSerAlaSerAla-----AsnGlnGlnAsnLys 2438
QY 8752 GATGAGTAAAGAAATTACAAACTTTACCTCTGCTCGAAGAAATGATTGATTAATAT 8811
Db 2439 AsnGly----- 2440
8812 CTGATATATTGGAAGCAGATACCTTCTTCTTAATAGTTCAGATGATTAACAAGGTATG 8871
2440 ----- 2440
QY 8872 ATTCTCCAGAGAGACATTTATGTACAAAGCATCATGCTGATATATTATAGAAA 8931
Db 2440 ----- 2440
QY 8932 GGTGATTAAGAAATTTTAAAAAAACTTCTTACTTCTGCTTTCAGTCAGCAATG 8991
Db 2440 ----- 2440
QY 8992 TTAGTCATAAATATAATGGAAGAAGAGTGTGCTTTCAGCAATGAATATAGTTAT 9051
Db 2440 ----- 2440
QY 9052 GCAGATTATCCGATATATTAAGAAGACATGATGACACATTCATATCTGA AAA 9111
Db 2441 ---GlnTyrLysAspAlaIle-----AspCysMetLeuLysLysLeuLysAspLys 2456
QY 9112 ATTAAAAAAATATTGGAACATCAATTAAGCAACGAAAAATCGTAAACATGTGGGAA 9171
Db 2457 Ile----- 2457
QY 9172 AATAATAGACGTCAGATATGCCACGCTATGTATGTGATATTAATTTGCTACTTCAAA 9231
Db 2457 ----- 2457
QY 9232 GTAACTATAGAGAGATGCTGTCAATTACCAAGATGAAGAACTAATCAGTTTCT 9291
2458 -----GlnGlnLys----- 2460
QY 9292 CGTTGTTAATTGAATGGCAAGCAAGCATGTAAAGAAAACATGTAAGTATCA 9351
Db 2461 -----GlnLysLysHisHisGlnThrSer 2468
QY 9352 TTA AAAACAATGT-----CCTCGTTCAAGCAAGATATATTTGAAGC 9396
2469 ---AspThrGlnCysSerAspThrProGlnProGlnThrLeuGlnAsp----- 2483
QY 9397 TCAGAAATTAATTAAGCAACCTGATGTCAAGATATTAAGAAATATATTAAGCTTGAAT 9456
2484 ---GlnThrLeu-----AspAspAspIleGlnThrGlnGlnLysLys 2497
QY 9457 ATATGATTAATAATACATGGAATGAAATGAAATATTAATTAATTAATTAATTAATTA 9516
2498 AsnMetMetProLysIleCysGlnLysVal---LeuLysThrAlaGlnGlnLysGln 2516
QY 9517 TCTTCAGGTATATATAGCAATTAACATGAGAGAAATGTTCACTCATATATAAATCA 9576
2517 ---GlnGlyCysValProAlaGlnLysSerGlnLysIleProAla-----Ala 2530
QY 9577 AAAGATTCTCAATGCGCTTGGAGTTAAATGATTAATTAATTAATTAATTAATTAATTA 9636
|||||
```

```
Db 2531 ThrAspSer----- 2533
QY 9637 AATAATGAATAATGAATTCAAAGAGTACTAAAGAAATTAATATCTGTTATATTT 9696
Db 2534 GlyLysGlnThrProGlnGlnThrProValLeuLys----- 2545
QY 9697 GTTGAAGATGAACACACAAATAATCATGTACTAGATGAATAATTAAGACAGACAA 9756
2546 -----ProGlnGlnGlnAla 2550
QY 9757 ACAGTTGCTCTAAAGCACTATTTCTTACACCCCATGATGATTTCTTATACAGA 9816
Db 2551 ValProGlnProPro-----ProProProProGlnGlnLysAlaProAla 2565
QY 9817 CCTTTATTCTCAACACATGAGTACGACAAATATGATCTTAAAAAT----- 9861
Db 2566 Pro-----IleProGlnProGlnProProProThrProThrGlnLeu 2579
QY 9862 -----GATTAATTAAGAGTATATCTGCTGTTATATGATCGGCTTACGT 9909
Db 2580 LeuAspAsnProHisValLeuThrAlaLeuValThrSerThrLeuAlaThrSerValGly 2599
QY 9910 TTTG-----ATAGCGCTTCATTCATGAAGAAATAATTAATTCGCTGTG--- 9954
Db 2600 IleGlyPheAlaThrPheThrTyrPheTyrLeuLysLysLysThrLysSerValGly 2619
QY 9955 GACTGTTCCGATATCTGAATATCCCGAAGAGAGTATGAATGCTTACGTTGAATCC 10014
Db 2620 AsnLeuPheGlnIleLeuGlnIleProLysSerAspTyrAspIleProThrLysLeuSer 2639
QY 10015 AAAAATAGTATACATACATATATGAAGTGTCCATTAAGGCAAAACATATATATAT 10074
Db 2640 ProAsnArgTyrIleProTyrThrSerGlyTyrAlaGlyLysArgTyrIleGlyLeu 2659
QY 10075 GAAGAGATACI-----AGTGAAGATGAAGATAAATATAGTGGACTTACTCC 10125
2660 GlnGlyAspSerGlyThrAspSerGlyTyrThrAspHisTyr----- 2673
QY 10126 TCTGATATTAATTCATCCGAAGATGATGAAGATGAAGATTAATTAATTAATTAATTA 10185
Db 2674 SerAspIleThrSerSerGlnSerGlnLysIleGlnLysLeuLysAspIleLysVal 2693
QY 10186 CCAGTACTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10245
Db 2694 ProGlySerProLysTyrLysThrLeuIleGlnValLeuGlnLysProSerGlyAsnAsn 2713
QY 10246 ATACCAAGTAT----- 10257
Db 2714 ThrThrAlaSerGlyAsnAsnThrThrAlaSerGlyAsnAsnThrThrAlaSerGlyLys 2733
QY 10258 GATACACCAAGT-----AATGATACACACGATGATGATGATTT 10296
Db 2734 AsnThrProSerAspThrGlnAsnAspIleGlnLysAsnAspGlyIleProSerSerLysIle 2753
QY 10297 ATTGATGATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 10347
2754 ThrAspAsnGlnTyrAspGlnLeuLysAspGlnIleThrLeuSerGlnTyrLeuGlnSerGln 2773
QY 10348 CCAATACAGAACCAAAATTAATTAACAAAAGTGCATATTCATTAATTAATTAATTAATTA 10407
2774 ProAsnThrGlnPro-----AsnMetLeuGlnTyrAsnValAspAsnAsnThrHisPro 2791
QY 10408 AATACCTTAATTTCTGATTAATCTGAGAGAAAACCTTTTATATATATATATATATG 10467
2792 ThrThrSerHis---HisAsnValGlnGlnLysProPheHisSerIleHisAspArg 2810
QY 10468 GATTATATATCTGGGAAGAAATTAATTAAT----- 10497
2811 AsnLeuPheSerGlyGlnGlnTyrAsnTyrAspMetPheAsnSerGlyAsnAsnProIle 2830
QY 10498 AATATTAATATGATTAAT-----ACTAATAATGATATATCAATG 10539
2831 AsnIleSerAspSerThrAsnSerMetAspSerLeuThrSerAsnAsnHisSerProTyr 2850
```

QY 10540 AAGCTAGAAATGATCTTATAGAGATATGATTAATGATTCAGTACGCTGCTA 10599
 DB 2851 AsnAspLysAsnAspLysSerGlyLeuAspLeuLeuAsnAspAlaLeu--SerGlyA 2870
 QY 10600 AACCTATATATATATGATGAGATATG 10628
 DB 2870 snHisIleAspIleTyrAspIleuLeu 2879
 RESULT 3
 AAY77905
 ID AAY77905 standard; Protein: 3060 AA.
 XX
 AC AAY77905;
 XX
 DT 13-JUN-2000 (first entry)
 XX
 DE Plasmodium var-7 polypeptide.
 XX
 KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
 XX DBP; Stialic Acid Binding Protein; SAMP; malaria; vaccine; Immunisation;
 XX Protoczoacide; var-7.
 XX
 OS Plasmodium sp.
 XX
 PN US5993827-A.
 XX
 PD 30-NOV-1999.
 XX
 PE 07-JUN-1995; 95US-0487826.
 XX
 PR 10-SEP-1993; 93US-0119677.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI SIm KL, Chitins C, Peterson DS, Su X, Wellem's TE, Miller LH;
 DR WPI: 2000-194198/17.
 DR N-PSDB: AA298287.
 XX
 PT Isolated protein binding domains from Plasmodium vivax and Plasmodium
 PT falciparum erythrocyte binding proteins useful for vaccinating against
 PT malaria -
 PS
 PS Disclosure; Columns 109-124; 93pp; English.
 XX
 CC The invention relates to ebl-1 polypeptides that are encoded by the DBL
 CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
 CC identical to the Duffy Antigen Binding Protein (DABP) and Stialic Acid
 CC Binding Protein (SABP), which are soluble proteins that appear in the
 CC culture supernatant after erythrocytes infected with malaria release
 CC merozoites. Immunochemical studies indicate that DABP and SABP are the
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
 CC and stialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
 CC used to vaccinate against malaria, especially caused by P. falciparum.
 CC Immunization with the polypeptide provides effective protection against
 CC malaria. The present sequence represents the var-7 polypeptide.
 CC
 XX
 SQ Sequence 3060 AA:
 Alignment Scores:
 Pred. No.: 5.39e-238 Length: 3060
 Score: 3501.00 Matches: 1057
 Percent Similarity: 38.65% Conservative: 470
 Best Local Similarity: 26.75% Mismatches: 934
 Query Match: 18.00% Indels: 1490
 DB: 21 Gaps: 147
 US-10-087-013-1 (1-10628) x AAY77905 (1-3060)
 QY 25 ATATATAAATGGGAGATGACATCATATAGAGGAGATGCTAAAGCCCTATTATA 84
 DB 4 LeuAlaLysMetGlyProLysGluAlaLeuGlyLysAspIleGluAsp----- 20

QY 85 AAGCAAAAGTCACAAAGTCAGCAAAATGTTTGGACGCTTATGCCAAATATAGACAT 144
 DB 21 -----GluSerAlaLysHisMetPheAspArgIleGlyLysAspValTyrAsp 36
 QY 145 CCATCAAAA---TATGCAAAAAGAACATGATGATTCGTTGCAAAAGCGATTTGACAAAGCA 201
 DB 37 LysValLysGluGluAlaLysGluArgGlyLysGlyLeuGlnGlyArgLeuSerGluAla 56
 QY 202 GAATTTGCGTGCGTCCCTTACGCGCAGTAAATAGCATATATATATATCCATATCA 261
 DB 57 LysPheGluLysAsnGlnLysSerAspProGlnThrProGluAsp-----Pro 71
 QY 262 TGTATATTAGATCATAGAGAACATATATATACGTTACGTTATGATGATGATGATGATGAT 321
 DB 72 CysAspLeuAspHisLysTyrHisThrAsnVal-----ThrThrAsnValIleAsn 88
 QY 322 CCTTGCACATGATGAGCAAAACCGCTTGTATGATGATGATGATGATGATGATGATGATGAT 378
 DB 89 ProCysAlaAspArgSerAspValArgPheSerAspLysGlyGlnCysThrHis 108
 QY 379 AATTAATATGATATATATTAAGCAAAAAGATGCTTACGCTGCGCCACCTAGAGA 438
 DB 109 AsnArgIleLysAspSerGlnGlnGlyAspAsnLysGlyLysAlaProTyrArgArg 128
 QY 439 CGACATATGTGTGATTAATAAAGCTGCAATGATATATATATATATATATATATATATAT 498
 DB 129 LeuHisValCysAspGlnAsnLeuGlnGlnIleGluProIleLysIleThrAsnThrHis 148
 QY 499 GATTATTTGGGAATGTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 558
 DB 149 AsnLeuLeuValAspValArgCysMetAlaLeuLysPheGlnGlnSerIleThrGlnAsp 168
 QY 559 CATCA---CATAAAGGACCT-----TCAGACGCTTACTGCTGCTGCA 600
 DB 169 TyrProLysTyrGlnAlaThrTyrGlyAspSerProSerGlnIleCysThrMetLeuAla 188
 QY 601 CGAAGTTTTCAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 DB 189 ArgSerPheAlaAspIleGlyAspIleValArgLysArgAspLeuTyrLeuGlyAsnPro 208
 QY 661 CATGAC-----AAGTAGAAAAGCGCTCGAGAGGTTTTCAGAAATA 705
 DB 209 GlnClnIleLysGlnArgGlnGlnLeuGlnAsnLeuLysThrIlePheGlyLysIle 228
 QY 706 CATGAT-----GGAATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 756
 DB 229 TyrGluLysLeuAsnGlyAlaGluAlaArgTyrGlyAsnAsp-----ProGlu----- 244
 QY 757 GGAATATATATATATATAGAGAACATGCTGATGATGATGATGATGATGATGATGATGAT 816
 DB 245 -----PhePheLysLeuArgGluAspTrpPheAlaAsnArgGluThrValTyrLys 262
 QY 817 GCTATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 876
 DB 263 AlaIleThrCysAsnAla---TrpGlyAsnThrTyrPhe---HisAlaThrCysAsnArg 280
 QY 877 CCATTATTTTCATATCTTAATGCGGCGCATTAACAGAAAGTTCCTTACCATTTTGAT 936
 DB 281 GlyLysArgThrLysGlyTyrCysArgCysAsnAspAspGlnValProThrTyrPheAsp 300
 QY 937 TATGTCCTTCATATTTTACGTTGCTGACGATGCGGAGAGAGATTTGCCGAAAAAGA 996
 DB 301 TyrValProGlnTyrLeuArgTyrPheGlnGluTyrPheLysAspPheCysArgGlyLys 320
 QY 997 AATATTAATTTGAAGAAAGCTGCAAGACTCGCTGCT-----AATGCAAAAGAA 1044
 DB 321 AsnLysLysIleLysAspValLysArgAsnCysArgGlyLysAspLysGluAspLysAsp 340
 QY 1045 CGCTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1104
 DB 341 Arg---TyrCysSerArgAsnGlyTyrAspCysGluLysThrLysArgAlaIleGlyLys 359

Db 1057 ThrAspMetClnThrArgLeuIleThrValPheLysAsnIleLysGluLysHisAspGly 1076
QY 2947 ---AAGAAATGATAATTAATTAATGATGAT-----GCCCCCAAAATTTAAATTTG 2994
Db 1077 IleLysAspAsnProLysTyrThrGlyAspGlnSerLysLysProAlaTyrLysLysLeu 1096
QY 2995 AGGAAATTTGGTGGAGAGCTAATAGACCAAAAGTATGGCAACCCATGAATGTATATA 3054
Db 1097 ArgAlaAspTyrPrpGlnAlaAsnArgHisGlnValTyrPrpAlaMetLysCys----- 1114
QY 3055 AATATTGTAAGATTAATCCGGACACCAATCAACAAAGTATGTTATGGCATATAGT 3114
Db 1115 -----AlaThrLysGlyLleIleCys----- 1121
QY 3115 GATCATACACCATTTGATGATATATATCCACAAAATTAAGATGATGACCAATGGCA 3174
Db 1122 ProGlyMetProValAspAspTyrLleProGlnArgLeuArgTyrPheThrGluTyrPheAla 1141
QY 3175 GAATGGTACTGCAGAGTGCAGAAAAAGAGTATGATAGTTGAAGAGAGAAGTGAAGAG 3234
Db 1142 GluTyrPtyCysLysAlaGlnSerGlnGluTyrAspLysLeuLysLysIleCysAlaAsp 1161
QY 3235 TGTAAAGATTAAGATTAATGTCAGAGCTGACAGAAAGAGTGTACAGTTGTACGAAG 3294
Db 1162 CysMetSerLysGlyAspGlyLys---CysThr---GlnGlyAspValAspCysGlyLys 1179
QY 3295 TGCACAGAAAGCTTGTATGATTAATTAATGATTAATGATGATTAATGATTAATGAT 3354
Db 1180 CysLysAlaLalaLysAspLysTyrLysGlnGluIleGlnLysTyrAsnGlnGlnTyrPheArg 1199
QY 3355 ATTAATTCAGATTAATTAACAAAGATTAATGATGATGATGATGATGATGATGATGAT 3414
Db 1200 LysIleSerAspLysTyrLysLeuLeuTyrLysGlnAlaLysThrThrSerThrAsnPro 1219
QY 3415 GGTATTGAAGCTCTAGTACTGCCAAAAATCATATATGAC-----AGGAATGTTAT 3465
Db 1220 Gly-----ArgThrValLeuGlyAspAspAspProAspTyrGlnGlnMetVal 1235
QY 3466 GAATTTTGTGGAAATTAATTAACAAATGCTGGCAAA-----AGTAATTA 3513
Db 1236 AspPheLeuThrProIleHisLysAlaSerLleAlaAlaArgValLeuValLysArgAla 1255
QY 3514 AGTGATACAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3573
Db 1256 AlaGlySerProThrGluIleAlaAlaAlaAlaProIleThrProLysSerThrAlaAla 1275
QY 3574 GCATATCTCCATGATACAGAAATTTGATGATGATGATGATGATGATGATGATGATGAT 3633
Db 1276 GlyTyrIleHisGlnIleGluIleGlyTyrGlyCysGlnGlnIleThrGlnPheCysGln 1295
QY 3634 GAAAA-----AGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3675
Db 1296 LysLysHisGlyLalaThrSerThrSerThrThrLysGlnLysLysGlnTyrThrPheLys 1315
QY 3676 GATTAACACACAGACCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3735
Db 1316 GlnProProProIleTyrLalaThrAlaCysAspCysIleAsnArgSerGlnThrGln--- 1334
QY 3736 GTACAGATTAATAAGAAAAAGGAGAAAGAAAGATGATGATGATGATGATGATGATGAT 3795
Db 1335 -----GlnProLysLysLysGlnGlnLysAlaGlnSerAlaCysLysLysLysLys 1351
QY 3796 GATTAATCTTAAGAAAAAGATGAGAAAGAAAGATGATGATGATGATGATGATGATGAT 3855
Db 1352 LysIleLeuGlnIleLysAsnGlyArgThrValGlyCysAsnProLysGlnSer 1371
QY 3856 AGTAATGATATGCGCATGCGCAATGC---GGAATATATAATTTAGTGAAGACCTCGT 3912
Db 1372 -----TyrProAspTyrPheCysLysAsnAsnIleAspIleSerHisAsp---Gly 1387
QY 3913 GTGTATGCGCCCTGAGAACAAAGTATGCTATGCTATGCTATGCTATGCTATGCTATGCTAT 3972
Db 1388 AlaCysMetProProArgGlnLysLysLeuCysLeuTyrTyrIleAlaHisGlnSerGln 1407

QY 3973 ATTAATAATTAATCAATCAAGTAAATTAATAAGAGCTTTCATCAATCTGCAGCAGCA 4032
Db 1408 ThrGlnAsnIleLysThrAspAsnLeuLysAspAlaPheLleLysThrAlaAlaAla 1427
QY 4033 GAACATTTCTTCATGATTAATTAATAAGTAAAG---GATGCTGAGGAATGATGATGAT 4089
Db 1428 GluThrPheLeuSerThrPrpGlnTyrLysSerLysAsnAspSerLysLysLysLysLys 1447
QY 4090 GATTAAGATTAATAAGCAAAATTCCTCCCGATTTTGTGATGATGATGATGATGATGATGAT 4149
Db 1448 AspArg-----GlyLeuIleProSerGlnPheLeuArgSerMetLeuTyrThr 1463
QY 4150 TTTGGAGATTAATGAGATTTTATTTTGAACAGATATATCAAAAGTGTATGCTGATGCTG 4209
Db 1464 PheGlyAspTyrThrArgAspLysLysLeuAsnThrAspIleSerLysLysGlnAsnAspVal 1483
QY 4210 ACTAAATTAAGAGCAAAATGATTTCTTTTCAAAAATGGTGACCAAAAATCTCTTAAT 4269
Db 1484 AlaLysAlaLysAspLysLysLysLysLysPhePheSerLysAspLysSerLysSerProSer 1503
QY 4270 GGAATAACAGCAAGAAATGCTGACAGAAACATAGTCATGATGATGATGATGATGATGAT 4329
Db 1504 GlyLeuSerArgGlnGlnIleTyrPtyLysThrAsnGlyProGluIleThrPtyLysGlyMetLeu 1523
QY 4330 TGTGACATGATTAATAATTTGGGCAAAAAGATGAT-----TTTACCGAAAACCTAC 4380
Db 1524 CysAlaLeuThrLysTyrValThrAspThrAspAsnLysArgLysLysLysAsnAspTyr 1543
QY 4381 GGTACACACAGCTCAAAATTTAGTGACAAA---AGCACACCTTTGAGAGAAATTTGCCAAA 4437
Db 1544 SerTyrAspLysValAsnGlnSerGlnAsnGlyAsnProSerLeuGlnIlePheAlaAla 1563
QY 4438 CGACCCAGATTTTATGATGCTTACACGATGATGATGATGATGATGATGATGATGATGATGAT 4497
Db 1564 LysProGlnPheLeuArgTyrPheTyrIleGluTyrPheGlnGlnIlePheCysAlaGlnArgGln 1583
QY 4498 AATATTGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4554
Db 1584 LysLysGlnLysIleLysLysAspAlaCysAsnGlnIleAlaSerThrGlnGlnCysAsn 1603
QY 4555 ACA-----GATGATTAAGAAATGCGAGGACTACCTTAATTAATTAATGAAA---AAA 4602
Db 1604 AspAlaLysHisArgCysAsnGlnAlaCysArgAlaTyrGlnGlnIleTyrValGlnAsnLys 1623
QY 4603 AAAAAAGAGTGTCCACAAGATAATTAATTAACAGAGTGAACGCAAAAAAAGATTC 4662
Db 1624 LysLysGlnPheSerGlyGlnThrAsnAsnPheValLeuLysAlaAsnValGlnProGln 1643
QY 4663 GATTAACAACACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4722
Db 1644 AspProGlnTyrLysGlyTyrGlnTyrLysAsp----- 1654
QY 4723 TTCAACAGCAAAATTTCTGCTAGTGTGCTGATTAACCTTGGAAGTGCCTCTGCTGATCAA 4782
Db 1655 -----GlyValGlnPro----- 1660
QY 4783 AGAATATATCAATTTGTAGAAAAAGCGTTACTATGATGCGCAGCAAACTGTGGTGC 4842
Db 1661 GlyAsnGlnTyrLeuLeuGlnLys-----CysAspAsnAsnLys---CysSerCys 1676
QY 4843 ACAAAATTTATGAAAAATGACAGCAAAATATACTTAACATTTGAGTAAGATGATGATGAT 4902
Db 1676 ----- 1676
QY 4903 GATTAAGTAAGAGGCAAAACACAGTGCTATTAAAGTGCACAAACAAAGTCTTAATAAC 4962
Db 1676 ----- 1676
QY 4963 TACAATACTTAAGAAATTAAGTGAAGATGCTTTTCTTCTGCTGCTGCTGCTGCTGCT 5022
Db 1676 ----- 1676

QY 5023 TGTTCATGCAATGGATGCAATTTATACAGATCCAGAAAGTTAAAGATCAAAATGGGTG 5082
Db 1677 -----MetaspGlyAsnValLeuSerValSerProLysGluLys----- 1689
QY 5083 CGAAAAAGATTGATGGAAGTGGCGCAACGAGGATTCATTTGGGTCATCTACTACAAA 5142
Db 1690 -----ProphegGlyLysTyr----- 1694
QY 5143 GAAAAAAGAAAAAGAAAAATATAAAACGTGGATGGCGCAAAATTTCTTAGAGGTC 5202
Db 1695 -----AlahsLysTyrPro----- 1699
QY 5203 CCGCCTTGCTAGTGTATGAAATATAGTTTATGATTTAAGAGATTAATTTAGGTATT 5262
Db 1699 ----- 1699
QY 5263 GATTAATTTGNAAGTGAAGAAAAACAAAGACGAGAAATTTGAAGAAATATTTAACAA 5322
Db 1699 ----- 1699
QY 5323 AATGGAACATCAGTTGGCAAAAGAGATGATGACTACAGAAATCCCGGTAGTACTGCG 5382
Db 1699 ----- 1699
QY 5383 CGAAATTTTTCGGAACGAAATTAAGGATGTGTGGAACCGCATGATATCGGGTAC 5442
Db 1699 ----- 1699
QY 5443 AAACGTGTAGGATGATGAAATAGTGAATATGTCAGAAAGTGTGAAGATCTAA 5502
Db 1699 ----- 1699
QY 5503 AAATGTGTCTGTACTTCAATGATGATTTCTTATGGGAAAAATCGCGATGAAGT 5562
Db 1699 ----- 1699
QY 5563 ACTGCGTATCAGTTCTGTGATGTTCGCGAATGGGGAAGATTTTTCGAACATAA 5622
Db 1699 ----- 1699
QY 5623 GAAAAAGAAATGGAATAATGGTAGGGCGGTAAATGATTAATCTGTGTATATGAA 5682
Db 1699 ----- 1699
QY 5683 GATTAAGAAGAAATGTACAGATGCGGTACACAAATATAAAATTTTACTGAGTGG 5742
Db 1700 -----GluLysCys----- 1702
QY 5743 AAACCAAGATGAAAAACAAATCAAAATATATGTGAGATAAAGACAAATATATTC 5802
Db 1702 ----- 1702
QY 5803 GAGCATCTGTGGCAAAAGATGACAGAGCGTCCGCAATATTAGACAAATTA 5862
Db 1702 ----- 1702
QY 5863 AAAATTTGTGAATAAAGTGAAGATTTGTAAATAGTATGAAGATGTCCACA 5922
Db 1703 -----AspCys----- 1704
QY 5923 CAGCATTAACATGATGATTAATAGTCAAAATATGCCGATCATTTAGAGATGAACAAA 5982
Db 1704 ----- 1704
QY 5983 GAAATGAAGAAATGTAATTTGCAAGTGCACAGAGTCCACACGCTGACGAAGGAA 6042
Db 1705 -----TyrGlnGlyLys-----HisValProSerLysPro----- 1715
QY 6043 ACACCGTACACAGGGGTATCACTGATATCAAAAGCGAGCATGAAAAAGAACGAA 6102
Db 1715 ----- 1715
QY 6103 ACAGGCGCGCTACAAAACAGCCGAAAAAGTGAATAATTAAACAGAAATGCGAGCA 6162
Db 1716 -----ProProProProValGlnProGln----- 1723
QY 6163 CAAACAGAACCCGAGAGAGCACAACAAACAGAAAGAAACATCAACAGCAACACA 6222
Db 1724 -----ProGlnAlaProThr----- 1728
QY 6223 ACAGAACTGACGTGGCGCAATGTAAAGCCATTCTTCGAATTAACACAGATCAGAG 6282
Db 1729 ValThrValAspValCysSerLysValLysThrLeuPheLys-----AspThrAsn 1745
QY 6283 GGTGAATAGAGGGTGTGATATCCAAAACGTATGACATAT-----CCTAAATGGCGT 6336
Db 1746 AsnPheSerAspAlaCysLysLeuLys-----TyrGlyLysThrAlaProSerSerThrLys 1764
QY 6337 TGTATT-----GTAGTAACTTAAGAAAT 6363
Db 1765 CysIleProSerAspThrLysSerGlyAlaGlyAlaThrThrGlyLysSerCysSerAsp 1784
QY 6364 GAAAAATGCGATATGATGCTGCTAGAGGAAAAATTAATGATTAATTAATACATAT 6423
Db 1785 SerGlySerLysCysIleProProAlaGlyAlaGlyValGlyLysLeuGln--- 1803
QY 6424 TTAATTAAGAACT-----GAAATTAACGCTGAC 6453
Db 1804 -----GluThrPalaThrAlaLeuProGlnGlyAlaAlaProSerHisSerThrAla 1822
QY 6454 AATGATATAAAGAGCGTTTATTAATGTCAGCAATGAACTCAATTTTGTGTTA 6513
Db 1823 AspAspLeuAlaGlyAsnAlaPheLysLeuSerAlaAlaIleGluThrPhePheLeuThrAsp 1842
QY 6514 AAATATATAATTAAGAA--AATCGTCAGACAGAAATGA-- 6549
Db 1843 ArgTyrLysGluGluLysLysProGlnLysPheLysSerGlnAlaLeuSerGlnLeu 1862
QY 6550 -----TTGCAAAATGGAACA 6564
Db 1863 ThrSerThrTyrSerAspAspGluLysAspProProAspLysLeuLeuGlnAsnGlyLys 1882
QY 6565 ATTCAGATGAATTTAAAGAAATATGATATATCATATGCTGATTAATAAGATGTTT 6624
Db 1883 IleProProAspPheLeuAlaGlyLeuMetPheTyrThrLeuGlyAspTyrArgAspIleLeu 1902
QY 6625 TTGGAACGTAAATTTCTAATGATTAATAAATAATTAATCTGATCAATATGATGAACAC 6684
Db 1903 ValHisGlyGlyAsnThrSerAspSerGlyAsnThrAsnGlnLysSerAsnAsnAsnIle 1922
QY 6685 ATTCATGAATGAATTAATAGAAAAAACAGATTAATAAAGATGAAGATTAACATAA 6744
Db 1923 ValLeuGlnAlaSerGlyAsnLysGluAspMetGlnLysIleGlnGlnLysIleGln 1942
QY 6745 ATA----- 6747
Db 1943 IleLeuProLysAsnGlyGlyThrProLeuValProLysSerSerAlaGlnThrProAsp 1962
QY 6748 -----TTTGGGAGAAAAATAAATAATTTATTTGGCAAGATGATATGATTAATCTAT 6804
Db 1963 LysThrPheAsnGlnHisAlaGlnSerLysThrLysGlyMetIleCysAlaLeuThrTyr 1982
QY 6805 -----CATCTCAGACGAAACGAAAAAGAAAAATTAACA 6840
Db 1983 ThrGluLysAsnProAspThrSerAlaArgGlyAspGluAsnLysIleGlnLysAspAsp 2002
QY 6841 GATTAATTAC----- 6849
Db 2003 GluValTyrGluLysPhePheGlySerThrAlaAspLysHisCysLysThrAlaSerThrPro 2022
QY 6850 -----CAGTACAT--GACATGACCAAACTG----- 6873
Db 2023 ThrGlyThrTyrLysThrGlnTyrAspTyrGlyLysValLysLeuGlnLysPheSerGly 2042
QY 6874 -----ACGCTTCCTTGAAGAGCTTTGTAATAAGCGCC 6906

Dh	2043	Ala	ysThr	Pro	Se	rAla	Se	rSer	Asp	Thr	Pro	Leu	Se	rAsp	Pro	Val	Leu	Arg	Pro	2062	
QY	6907	CA	ATTTTG	AGATG	CTTC	ACAC	AA	TG	GC	AG	AA	ATTTT	GT	TAAT	TA	AG	AG	AG	AC	6966	
Dh	2063	Pro	Thr	Phe	Arg	Thr	Leu	Arg	Leu	Arg	Leu	Arg	Leu	Arg	Leu	Arg	Leu	Arg	Leu	2082	
QY	6967	TTG	TTAA	AAAT	TGG	AGG	GGC	CG	TAA	GA	AA	TTG	AG							7002	
Dh	2083	Leu	Ala	Gln	Leu	Leu	His	His	Glu	Cys	Leu	Val	Glu	Glu	Asn	Gly	Gly	Ser	Arg	2102	
QY	7003	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2122	
Dh	2103	Gly	Leu	Thr	Arg	Gln	Thr	Ser	Gly	Asp	Gly	Ala	Cys	Asn	Glu	Met	Leu	Pro	Lys	2142	
QY	7018	G	AGCG	TAA	G	A	C	A	G	A	A	-----	-----	-----	-----	-----	-----	-----	-----	7062	
Dh	2123	Asp	Leu	Thr	Val	Pro	Asp	Leu	Arg	Leu	Arg	Leu	Arg	Leu	Arg	Leu	Arg	Leu	Arg	2142	
QY	7063	A	ATTTT	TAA	GA	AG	TGG	AAA	ACT	GAA	AT	TGA	AA	AC	CA	AA	AG	AA	AA	7122	
Dh	2143	Lys	Thr	Leu	Ser	Leu	Ser	Gly	Ser	Glu	Phe	Glu	Gly	Ser	Glu	Gly	Ser	Glu	Gly	2162	
QY	7123	AA	AG	AT	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	7170	
Dh	2163	Lys	Asp	Pro	Cys	Val	Asn	Gly	Ser	Asn	Lys	His	Asp							2174	
QY	7171	G	A	G	A	G	C	A	A	C	A	T	G	C	T	C	A	T	A	7230	
Dh	2175	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2175	
QY	7231	A	AG	AT	T	G	T	C	T	G	T	A	G	C	A	A	A	A	A	A	7290
Dh	2176	Gly	Phe	Cys	Glu	Thr	Leu													2182	
QY	7291	CA	AT	CC	AG	T	CC	T	A	T	G	AT	T	GC	AG	A	T	GC	T	7350	
Dh	2183	Thr	Ser	Thr	Val	Ala	Lys	Asp	Phe	Leu	Thr	Leu	Gly							2195	
QY	7351	A	AG	T	G	A	G	T	C	C	T	G	A	A	A	A	A	A	A	A	7392
Dh	2196	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2213	
QY	7393	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	7440	
Dh	2214	Phe	Lys	His	Thr	Lys	Asp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2230	
QY	7441	A	A	G	C	A	C	A	T	A	T	T	A	T	T	A	T	C	T	A	7500
Dh	2230	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2230	
QY	7501	A	A	A	T	T	A	C	C	T	A	T	T	A	G	T	C	T	A	A	7560
Dh	2231	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2238	
QY	7561	A	A	T	C	T	G	C	A	T	C	T	A	A	C	T	A	T	A	T	7620
Dh	2239	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2242	
QY	7621	T	G	T	A	A	A	T	G	A	A	A	A	A	A	T	C	G	T	T	7680
Dh	2243	Cys	Arg	Asn	</																

Oy	7858	-----GAGAAACGGGTCCGAATGCAATCCAAATATGTGAATCATATGAATATATAGTTTGCCT	7911
Dd	2279	GlyLeuGIuAsnAlaCys	2284
Oy	7912	GATCTGGGTGCATAGTTTAGAGAACAGATATGTATTACGAATTTGGTTACTTAACCTCCC	7912
Dd	2285	-----ArgGLylAlaGly-	2288
Oy	7972	GTACAATAAATATTATTAAGGTTTTTGAAATCATATATATGCAAAAATCGAATAATAAAAAT	8031
Dd	2289	-----IlePheGLuGlyIle-	2293
Oy	8032	AAAGGTGAATAAATATACAAACGATGTACAAACGTTTCGTTCTGTGGGATGCTAAT	8091
Dd	2293	-----	2293
Oy	8092	AGAAAGAATATTGGAAAACA-----ATGAGCTGCMAAGCACACGAA	8133
Dd	2294	ArgLysAspGIuITrPlysCysArGAsnValCysGLyTyValValaICyLys---ProIIu	2312
Oy	8134	GATGCCAAACCTTTTACAAAAGGAAGAATGATGAGATTGGAACGCATACNTAATACAA	8193
Dd	2313	AsnValAsnGLyGlualalALysGLyLys-----	2321
Oy	8194	GATTAAGTGTGCATTAAGACAGATCCACCCTGTGATTAATTAATACCTCAACGGTTTCCA	8253
Dd	2322	-----HisIleIlle6InIleArgIla	2328
Oy	8254	TGGATGACTGAATGGCTGAATTTATTTATGTAAAGCACTGATGGAAGAAATTGGAAAAATT	8313
Dd	2329	LeuValLIysAsrGItPrValGIuTyPrHe-----PheGLiasPIryTrsnLysIle	2344
Oy	8314	AAAAAATCATGTGTGATCCTGTAAACATCTGCACAGATGCAGAAATGATATGATGAAAT	8373
Dd	2345	LysHIs-----Gly	2355
Oy	8374	AAGGTCAACAGTGTAAACAGAGATGTCAAAGATTAATAAATTTGTTCTTAATGCAAA	8433
Dd	2356	GIuILeserProcys-----IleLysAsnCysValGIuILystr--	2368
Oy	8434	TCTCTATCGATATACATCAATTAATAATACAAAGATTTGTAGCAACAATATATACA	8493
Dd	2369	-ValAspGIuLysArGLySGluITrPlysGIuIleThGIuArg--	2382
Oy	8494	AAAAATCTCTACTTAATGATCATGTTCAAAATTTGTACAAAAGTTGAAAACTTTTAAAGT	8553
Dd	2383	-----PheLysAsp	2385
Oy	8554	GAATGTTCTGTGTGACCTTTTCTGAAATATCTTCATGAAACAAGTAAGTGTGAAATAT	8613
Dd	2386	GIu-tyr	2387
Oy	8614	AAATTTATGAATAATGATGGTCTTCCCAATATACGAACATATGCTTTCGAAAGAAACACA	8673
Dd	2388	Lys---AsnAspAsnserAspAspAsnValArgSerPheLeuGIuThIleuIlePro	2406
Oy	8674	AAA-----AGTTATPAAGACCT	8691
Dd	2407	GIuIleThrAspAlaAsnAlaLysAsnLysValIleLysLeuSerLysPheGLyAsnSer	2426
Oy	8692	TGCAGTTGTACACTACTCTTAAGAAATTCATGATTAATTTGCTACCGATCAACAAACAA	8751
Dd	2427	CysGLyCysSerAlaSerAla-----AsnGIuGIuAsnLys	2438
Oy	8752	GATGAGATGAAGAATTAACAACCTTTTACCTCTGCTCGAAGAAATGATTAATTAATAT	8811
Dd	2439	AsnGLy-----	2440
Oy	8812	CTTATATTAATTTGAAACGATACCTTGTTTCTTAATAGTTTCAGATGAATACAAAGGTGTATGG	8871
Dd	2440	-----	2440


```

OY 8872 ATCTCCAGAAGAGACATTATGTACAGACCTATCATATATATATAGAAAA 8931
DB 2440 ----- 2440
OY 8932 GGTGATTAAGAAATTTTAAAAAAACCTTCCTACTCTGCTTTCAGTCAGACAAATG 8991
DB 2440 ----- 2440
OY 8992 TTAGGTCAAAAAATATATAATCGAAGAGAGTGTGCTTTGAGCAATGAATATAGTAT 9051
DB 2440 ----- 2440
OY 9052 GCAGATTAATCCGATATATTAAGAACTGATATGATGACACTTATATCTGAAAA 9111
DB 2441 ---GTTTTLysAspLalle-----AspCysMetLeuLysLeuLysAspLys 2456
OY 9112 ATTAATAAAATTTTGAACAATCAATGAAGAACCGAAAAATCGTAAAAACATGGCGGA 9171
DB 2457 Ile----- 2457
OY 9172 AAATATAGACGTCAATATGCGACCGCTATGTATGTGATATATAATGCTACTCAAA 9231
DB 2457 ----- 2457
OY 9232 GTAACATTAGATGAAGATGGTGTCAATTACCAAGAGTGAAGAACTATCAGTTTCT 9291
DB 2458 -----GlyClnCys----- 2460
OY 9292 CGTTGTTAATGTGAMGGCAAGCAGCAGTGAAGAAAGAAAGATAGTATCA 9351
DB 2461 -----GlnLysLysHisHisLysGlnThrSer 2468
OY 9352 TTTAAACAAATATG-----CCTCGTTCAACAGAGATATTTTGAAGCG 9396
DB 2469 ---AspThrGlnCysSerAspThrProGlnProGlnThrLeuGlnAsp----- 2483
OY 9397 TCAGATTAATTAGACACACCTGATGTCAATGATTAAGAAATATATAGCTTGAAT 9456
DB 2484 ---GlnThrLeu-----AspAspAspIleGlnThrGlnGlnLysLys 2497
OY 9457 ATATGATTAATAATATACATGAAATCTAATATATAATATAGCAATTAAGATCA 9516
DB 2488 AsnMetMetProLysIleCysGlnAsnVal-----LeuLysThrAlaGlnGlnLysGln 2516
OY 9517 TCTTCAGGTAATATAGACAAATACACCTGAGAAAGAAATGTTCACTATATTAATCA 9576
DB 2517 ---GlyGlnCysValProAlaGlnLysSerGlnGlnProAla-----Ala 2530
OY 9577 AAAGATTCATGCGGCTTTGGAGTTAATGATATTAATGAATAGTTACAGAACAAAA 9636
DB 2531 ThrAspSer----- 2533
OY 9637 AATATGAAATATATGAATTCACAAAGAGTACTAAAAAATATATCTGTTATATTTT 9696
DB 2534 GlyLysGlnThrProGlnGlnThrProValLeuLys----- 2545
OY 9697 GTTGAAGATGAACACACAAAAATCATGACTAGATGAATATTAAGAAAGACAGCA 9756
DB 2546 -----ProGlnGlnLysLys 2550
OY 9757 ACAAGTCTGCTAAAGCACTATTTCTTTACACCCCATGATGATCTTTCTATACAGCA 9816
DB 2551 ValProGlnProPro-----ProProProProGlnGlnLysAlaProAla 2565
OY 9817 CTTTATATCTGAACACATGAGTAGCACAATATGATCCTAAAAAT----- 9861
DB 2566 Pro-----IleProGlnProGlnProProThrProProThrGlnLeu 2579
OY 9862 -----GATATATGCAAAAGTAGTAGTCTGTTGTTATGTATCGCGCTTAGGT 9909
DB 2580 LeuAspAsnProHisValLeuThrAlaLeuValThrSerThrLeuAlaThrSerValGly 2599
OY 9910 TTG-----ATAGCGCTTATTCATGAAAGAAAAATTCAAATCGTCTGTG--- 9954

```

```

DB 2600 IleGlyPheAlaThrPheThrTyrPheThrTyrLeuLysLysLysThrLysSerValGly 2619
OY 9955 GACTGTGGCCGATATACGAATATCCCGCAGAGAGCTATGCAATCCGTCAGTGAATCC 10014
DB 2620 AsnLeuPheGlnIleLeuGlnIleProLysSerAspTyrAspIleProThrLysLeuSer 2639
OY 10015 AAAATAGTACATACATATATGAAGTGTCCATTAAGCGCAAAAACATATATATATG 10074
DB 2640 ProAsnArgTyrIleProTyrThrSerGlyLysTyrArgGlyLysArgTyrIleTyrLeu 2659
OY 10075 GAAGAGATFAC-----AGTGAGATGAAGATTAATATATGTGGACTTATCTTCC 10125
DB 2660 GlnGlyAspSerGlyThrAspSerGlyTyrThrAspHisTyr----- 2673
OY 10126 TCTGATTAATCTTCATCCGAAGGTAGATAGAGATGAGATTAATATATATATATGTA 10185
DB 2674 SerAspIleThrSerSerGlnSerGlnSerGlnSerGlnSerGlnSerGlnSerGlnSer 2693
OY 10186 CCAGGTAGTCTTAATATATAAATCATGATAGAGTACTAGAACATCAAAAAAGGAT 10245
DB 2694 ProGlnSerProLysTyrLysThrLeuIleGlnValIleLeuGlnProSerClyAsnAsn 2713
OY 10246 ATCCAGATGAT----- 10257
DB 2714 ThrThrAlaSerGlyAsnAsnThrThrAlaSerGlyAsnAsnThrThrAlaSerGlyLys 2733
OY 10258 GATACACCAAGT-----AATGATACACCCGTCAGCAATATGATTT 10296
DB 2734 AsnThrProSerAspThrGlnLysAsnAspIleGlnLysAsnAspGlyIleProSerSerLys 2753
OY 10297 ATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10347
DB 2754 ThrAspAsnGlnThrPAsnGlnLeuLysAspGlnPheIleSerGlnTyrLeuGlnSerGln 2773
OY 10348 CCAATATACAGAACCAATATATATATTAACAAGTGCAGATATTTCCATGATATACAGAACCT 10407
DB 2774 ProAsnThrGlnPro-----AsnMetLeuGlyTyrAsnValAspAsnAsnThrHisPro 2791
OY 10408 AATCTTATATATCTGATTAATCCCGAAGAAACCTTTATATATCATATCATGATGAG 10467
DB 2792 ThrThrSerHis-----HisAsnValGlnGlnLysProPheIleMetSerIleHisAspArg 2810
OY 10468 GATTATATATCTGGGAAAGAAATAGTAT----- 10497
DB 2811 AsnLeuPheSerGlyGlnGlnTyrAsnTyrAspMetPheAsnSerGlyAsnAsnProIle 2830
OY 10498 AATATATATATGAGTACTAAT-----ACTAATATATGATATTTCAATG 10539
DB 2831 AsnIleSerAspSerThrAsnSerMetAspSerLeuThrSerAsnAsnHisSerProTyr 2850
OY 10540 AATGCTAGAAATGATCTTATATGAGGTATATGATTAATTAATGATGATGATGATGATGAT 10599
DB 2851 AsnAspLysAsnAspLeuTyrSerGlyIleAspLeuIleAsnAspAlaLeu-----SerGly 2870
OY 10600 AACCTATTGATATATATGATGAGATATG 10628
DB 2870 snHisIleAspIleTyrAspGlnMetLeu 2879

```

RESULT 4

AAW00384 standard; Protein: 2913 AA.

AAW00384;

21-FEB-1997 (first entry)

Plasmodium falciparum erythrocyte membrane protein.

Plasmodium falciparum; erythrocyte membrane protein; malaria; detection; identification; treatment; prevention; parasite.

Plasmodium falciparum MC type.

XX	Key	Location/Qualifiers
PH	Domain	62..394
FT	Domain	/label= Duffy binding ligand domain 1
FT	Region	607..648
FT	Region	/note= "Cysteine rich motif"
FT	Domain	839..1272
FT	Region	/label= Duffy binding ligand domain 2
FT	Region	1487..1527
FT	Domain	/note= "Cysteine rich motif"
FT	Domain	1706..2005
FT	Domain	/label= Duffy binding ligand 3
FT	Domain	2102..2349
FT	Region	/label= Duffy binding ligand 4
FT	Region	2354..2398
FT	Domain	/note= "Cysteine rich motif"
FT	Domain	2450..2475
FT	Domain	/note= "Putative transmembrane domain"
NN	MO9633736-A1.	
XX	31-OCT-1996.	
PF	26-APR-1996;	96MO-US05798.
XX	27-APR-1995;	95US-0430908.
XX	(AFYFV) AFYMAX TECHNOLOGIES NV.	
PA	Baruch DI, Howard RJ, Pasloske BL;	
XX	WPI; 1996-497376/49.	
DR	N-PSDB; NAT41852.	
XX	New Plasmodium falciparum erythrocyte membrane proteins - used to develop products for the diagnosis, treatment or prevention of malaria parasite infections	
PT	Claim 1; Figure 12; 149pp; English.	
PT	A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte membrane protein 1 (PFEMP1) or active fragments or analogues of that protein can be used in the treatment or prevention of symptoms of a malaria parasite infection. The polypeptides can inhibit, block or reverse the sequestration of erythrocytes in patients suffering from malaria. Nucleic acids derived from the PFEMP1 gene can be used as probes and primers to identify a Plasmodium falciparum parasite, the primers used to generate characteristic amplification patterns from different P. falciparum strains. Antibodies specifically immunoreactive with the PFEMP1 polypeptide or its fragments may be used in diagnosis of malaria infection. This is the PFEMP1 protein of the MC type of Plasmodium falciparum. An alternative, truncated PFEMP1 protein is given in AAM00385.	
XX	Sequence	2913 AA;
CC	Alignment Scores:	
CC	Pred. No.:	9.53e-212
CC	Score:	3128.50
CC	Percent Similarity:	38.47%
CC	Best Local Similarity:	26.49%
CC	Query Match:	16.08%
CC	DB:	17
CC	US-10-087-013-1 (1-10628) x AAM00384 (1-2913)	
CC	97 AAAAGTCGAGAAATGTTTGGACGTTATCCAAATAATA-----	138
CC	111 : : : : :	
CC	13 LysAspAlaIaLysSHsAlaLeuAspArgIleGIyIuGIuValYrYIySGIuYsValGIu	32
CC	139 AGACATTCATCAAAATATGACAAAGAACACATGTGGATTCGTTGAAGGCAATTGCACAAA	198
CC	33 AsnAspAlaGIuYrYIySGIuYs-----AlaLeuYsGIuYsAsnLeuGIuGIu	48

[illegible]

OY 1213 AAAGAAATACAA-----TCATATTATGCAAGCATTAACAATTTGCAATAT 1260
 :::||||| ::: ||| :::|||||
Db 396 GluGluIleAsnLysThrHisGlyThrThrIleThrThrGlyAsnGlyLysIleAsnAsn 415
OY 1261 ATTAATAGTAATTTATTAACATTTTATGAAAATTTAAGAAACGCAATATGCAACT 1320
 ::: ||| ::: ||| |||
Db 416 Leu-----TyrValGlyHisPheThrLysIleLeuLysLys---TyrTyrProThr 431
OY 1321 AATGACCTTTTAAATTTTAAATGAAAGAAAGTATGTTAA-----GGAGATT 1374
 ||| ||| ||| ||| |||
Db 432 ValAspLysSerLeuGlnLysLeuAsnAspGluAlaIleCysLysLysProProAsnVal 451
OY 1375 CCAGAGAAAAGCATATTAATCTTTACTTAACAGTCGTGATGCAAAAGGATATTTATCGT 1434
 ||||| ||| ::: ::: |||
Db 452 GlyAsnGlyLysAlaSerThrValAspPheAsnAsnGluValAsnThrThrPheSerHis 471
OY 1435 TCAGAAATATTTCCAGATGTGTCGCGACTGGGGGTCAAAATGTAT-----GGT 1482
 ::: ||||| ||| ::: |||
Db 472 ThrThrTyrCysGluAlaCysProTyrCysGlyAlaGlnLysGluLysAsnGlyGly 491
OY 1483 ATAAATATACACACAAATCA--GATATGATCGTGACGTGTAATATATGAAAGCAT 1539
 ||| ||||| ::: ||| |||
Db 492 TrpLysAlaLysGlySerCysAlaLysLysLysGlnArgIlePheAsnLysGlu--- 510
OY 1540 AAACCTCCATGGGGTGAAGCCTACTAATATGACTGTCTTATATGCTGAATGAACA 1599
 ||| ||| ||| ||| |||
Db 511 -----AsnSerThrAspIleLysIleLeuThrProGluLysGlyArg 524
OY 1600 GGTGATATTACACAAAATTTGAAATTTTGTACAGCTCAACTATTTACAAAGTATA 1659
 ::: ||||| ||| ||| |||
Db 525 SerLysThrLeuGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 542
OY 1660 AATATCAAAAAAGGAAATCTATTTAAGATGAAATATTAATATGATGTAACGTA 1719
 ||| ||| ||| ||| |||
Db 543 ---AsnAspIleTrpLysCysHisLysTyrAspAspAsnGly-----ThrAsp 556
OY 1720 CAAATACTGCAATCAATATGAT-----AAT 1746
 |||::: |||::: |||
Db 557 AspGlnThrAspAspSerAsnAspCysValLeuGlyAspTrpGlyAsnLeuThrLysGlu 576
OY 1747 CCAATATATATCATTTTATTAATTTTGTGATTTGGTATACATATTTATTAAGGAT 1806
 |||::: |||::: |||::: |||
Db 577 AspLysIleMetSerLysTrpAsnAlaPhePheTrpMetTrpValHisAspPheLeuLeuLeu 596
OY 1807 ACTATTAATGTAAGTACAAACTTAAACTGTATTAAT---AATACAAACCGCATTTGT 1863
 :::||||| |||::: |||
Db 597 SerIleLysTrpArgAspGlnLysGlyArgCysIleAsnLysAspLysGlyLysThrCys 616
OY 1864 ATTGATGAATGTACAGAAATGCTTATGTTTGAACAGATGGTAAACAAAGAAAGAA 1923
 ||| |||::: |||::: |||::: |||
Db 617 IleLysGlyCysAsnLysCysIleCysPheGlnLysTrpValGlnGlnLysLysThr 636
OY 1924 GAATGGAATAGTATAAGAAAGTGTTCACAAAAGAAAGAAATATACAGCATCGATTTAT 1983
 ||||| ||| |||::: |||
Db 637 GluTrpGlyLysIleLysAspHisPheArgLysGlnLysAspIleProLysAspTrp--- 655
OY 1984 AGTAATATTATTAATCTTTTGAAGTTATTTTAAAGTTATGATTAACCTTGACAA 2043
 |||::: |||::: |||
Db 655 ----- 655
OY 2044 GATGAAGCAAAATGCAAGAACTTATGCAAAATATTAAGAAAAAATAGAGTTTCC 2103
 |||::: |||::: |||
Db 655 ----- 655
OY 2104 AATTTGAAATTAATAGGAGCTATTTAGAGATGCAATA-----GAACCTTGTGTAGAT 2157
 ::: |||::: |||::: |||
Db 656 -----ThrHisAspAspPheLeuGlnThrLeuLeuMetLysAspLeuLeuGln 672
OY 2158 CACTTAAGAAAGTGCACAGATATGTAAAGACATATATACAAAGCAATGTGAACA 2217
 ::: |||::: |||
Db 673 IleIleGlnAspThrTyr-----GlyAspAlaAsnGlu----- 683
OY 2218 TCCCATATGCAACAAACCCGTCGTGTTAAACCTGCGAGGACGCAACCCGCACTAAA 2277

Db 683 ----- 683
OY 2278 AATATAAGAAATAGCACAAATACTTTAAAGAGTGCATAGAGAAAGCAAGAAATGCT 2337
 ||||| ||| |||::: |||
Db 684 ---IleLysArgIle-----GluAlaLeuLeuGlnGluAla----- 694
OY 2338 GGTCTTCATTAATTTGAAAGAAAGCAGCAGCAAGGTATATATTAACGTGGGGTAGAGA 2397
 |||::: ||| |||::: |||
Db 695 GlyValGlyGlyIleAspPheAlaIleLeuAlaGlyLeuTyrThrLysGlyPheValAla 714
OY 2398 AAGGACTTCAAGCAATTTATGTAATATATGATATAAACAATTTAAATCGTAATCTTGGT 2457
 |||::: |||::: |||::: |||
Db 715 Glu-----LysAspThrThrIleAspLysLeuGlnHisGluGlnLys----- 729
OY 2458 TTTCAATAGCACATGTGATGCGAAAGGACAGAGGTATGATATCAAAAGATTTGTC 2517
 |||::: |||::: |||
Db 729 ----- 729
OY 2518 GTAGCACTGAATGGAAAGTGCATCCGACACATGCTAAAGATCAGCAAGATGTATT 2577
 ::: ||| ||| |||
Db 730 -----GluAlaAspLysCysLeuLysThrHisThrAspAspThr 742
OY 2578 ATGCTCTAGAGAGACAGACATATATGTACATCCATTTGGAACATTTTCAACAGGAT 2637
 |||::: |||::: |||
Db 743 CysProProGln-----GluAsp 748
OY 2638 CACCACTTAATGTATATTTGTGATGATTTAGTTAATAATCTTTTGGGGATGTT 2697
 ::: |||
Db 749 ArgSerAla----- 751
OY 2698 CTTTATACGCAAAATATGAAAGCAACAAAGATATATAGCAATGTATAAGAAAGAAATAC 2757
 |||::: |||::: |||
Db 752 ---AlaArgSerGluSerAlaThrVal----- 759
OY 2758 CTAAGGGCCCCAAGAAAGTAACTGAACCCAAACAGACAACTATGTGCGAGTATA 2817
 ||| |||::: |||
Db 760 ---ProSerProProAlaAspProLys----- 767
OY 2818 CGTTACAGTTTGGAGATATATAGTGATATATTTGAGAGAAAGAGCTCTGGGAAAGAAC 2877
 |||::: |||::: |||
Db 767 ----- 767
OY 2878 GGTGACATGTTAAAGCTGCAAGACATTTGGAACCTTTTGGTAAATATACATTAAGTCA 2937
 |||::: |||::: |||
Db 768 -----AlaThrGluGluValAspAlaAsn 775
OY 2938 CTCAAAGCAAGCAATATGATTAATATATATGATGATGCCCCCAAAATATTTAAATGAGG 2997
 ||| ||| |||
Db 776 AlaSerSerAspGlyGluAspAspPheGluGluGlu----- 787
OY 2998 GAAATTTGTGGAGAGCTAATATAGAGCAAAAGTATGGGAACCATGAATGTATATAAA 3057
 |||::: |||::: |||
Db 787 ----- 787
OY 3058 TATTTGAGAGTAATTCGGACACCAATCAACAAAGTAGTTATTCGCGATATAGTAT 3117
 |||::: |||::: |||
Db 787 ----- 787
OY 3118 CATACACATTGGATGATTTATCCCACAAAATTTAAGATGATGACCCAAATGGCAGAA 3177
 |||::: |||::: |||
Db 787 ----- 787
OY 3178 TGCTACTGCAAGGTGAGAAAAGAGATAGTAAATGTAAGAGAAAGTAAAGAGTGT 3237
 ::: |||::: |||::: |||
Db 788 -----GluGluGluGluGluAspGluGlyGluGluGluAlaGluGluVal 802
OY 3238 AAGGATTAAGATATATGCTCAAGCTTACGAAA-----GAGAGTGTACAGCTTGT 3288
 ::: |||::: |||
Db 803 GlnGluGluLysThrAspGlnSerAlaThrGluAlaValAlaProSerProProGlyThr 822
OY 3289 ACGAAGTGCACAGAAAGCTTGTATATGAAATATATATATATATAGATTTATGCAAAAGACAA 3348
 |||::: |||::: |||

Db 823 ThrGln----- 824
QY 3349 TGCAGATATATATACAGTAATAATCAAGAAATACATGACACAGCAAAATGCTGTAGT 3408
Db 825 -----AspGlyValIlyProAlaSerGlnIlyAspAspValIlyValCys 839
QY 3409 AATAGTGTATGTAGAGTTCACAGTACGCGCAAAATCATATACAGAGAAATGTTATTGAA 3468
Db 840 SerIle-----ValAspLysAlaLeu----- 846
QY 3469 TTTTGTGCGAATATATACCAACAAATGTGCGAAAGTAATAAAGTGGTACTAGTAT 3528
Db 847 -----LysGlyLysLeuAsp 851
QY 3529 GAAAGTCTGTCAAT-----GCTATACACACACAGTATGAAATGTTGGAGCA 3576
Db 852 AspAlaCysThrLeuLysTyGlyLysGlyAspAsnThrThrGlnSerIle----- 869
QY 3577 TATCTCATGATACAGCAAAATTTTGATGATGTGCATCACAATAATGATTTGTGATGAA 3636
Db 869 ----- 869
QY 3637 AAAAGTATGTATAGATTAACGAAATATATGCTTAGAGATTAACCAACAGCAACATGAT 3696
Db 870 -----ThrLysPro----- 872
QY 3697 GGTGCGTGTGTGTTAAAGTGTGATCGCAACACAGAGGTACAGATAAAAAGCAAAAAA 3756
Db 873 GlyAlaIleGlyThrProSerGly----- 880
QY 3757 AAAGCGCAAGAAAAGTATCGCAATGTAAACAGTGAATATACTTAAAGAAACGAT 3816
Db 881 -----LysAspThr----- 883
QY 3817 GGAAGAAACAAGTAGAAGATGTGCATCAAAAAAGATAGATATCCCGATTTG 3876
Db 883 ----- 883
QY 3877 CAATGCGCAATATTAATTTAGTGGAGACCCCTGTGTATGCCCTTAGACAGCA 3936
Db 884 -----GlySerIle-----CysValProProAlaGlyArg 893
QY 3937 AAGTATGCTTACATTTCTTG----- 3957
Db 894 LysLeuValIleLysLeuHisAspTrpAlaGlyLysIleThrGlnAlaLysSer 913
QY 3958 -----GCAATGATTAAT----- 3969
QY 914 GlnGluThrSerGlyLysGlnLysThrProSerGlyAsnGlnSerProSerGlyLys 933
QY 3970 -----GAAATATAAAAAATTAACATCACAAGTAAATTTAAAAAGAA 4008
Db 934 LeuProGlnIlyProThrProGlnIlyThrLysGlnIlyThrProGlnIlySerLeuLeuHis 953
QY 4009 GCTTTCATCAAACTGACAGACAGCAAACTTCTTCATGATTAATTTAAAGTAAG 4068
Db 954 AlaPheValSerProProAlaGlyArgPheLeuProThrLysPheLysGln 973
QY 4069 -----GAT 4071
Db 974 TyrPheAlaGlnHisGlyAlaGlyAlaThrGlyGlnIleThrIleGlyThrLeuAsp 993
QY 4072 GGTCAAGCAAAATGAACTC-----GATAAAGAAATTAAGAGCAAAATTCCTCCCGCATTT 4128
Db 994 GlyLysIlyLysGlnIlyThrProAspLysLeuLeuLysThrGlnHisIleProProAspPhe 1013
QY 4129 TTGAGATCCATGTTCTACATTTGGAGATTATAGAGATTTTATTATTTGCA-----ACAGAT 4185
Db 1014 LeuAlaTrpIleMetPheTyThrLeuGlyAspTrpAlaGlyAspIleLeuValIleLysThrAsp 1033
QY 4186 ATA-----TCAAAAGTCATGTGAGCGAAGTAACAAAGCAAAATGATATCT 4236
Db 1034 IleValValHisThrSerGlyAsnLysGlyAspMetGlnIleMetGlnAlaIleGlnLys 1053

QY 4237 -----CTTTCAAAAAATGTGACCAAAAAATCTCTAATGCAAAAAACA--- 4278
Db 1054 LysIleGlnGlnIleLeuProThrSerGlySerSerProSerProProAlaValThrGln 1073
QY 4279 -----CGCAAGAAATGTGACACACAACTACTCATGAGATA 4314
Db 1074 ThrGlnHisSerValGlnAsnProAlaGlyThrTrpThrAsnGlnLysLysIle 1093
QY 4315 TGGAGAGCTTGTCTATGCTACTAGTA----- 4341
Db 1094 TrpGlnIlyMetValCysAlaIleLeuThrTyrAsnThrAspThrProSerGlyThrAlaPro 1113
QY 4342 -----AAAAATGGGCAAAAAA---GATGATTTTACCGAAMAC----- 4377
Db 1114 ThrGlnIleGlnIleValAlaGlyThrLysLeuThrAspLeuAsnSerLysAsnProLysIle 1133
QY 4378 -----TACGTTACACAAACAGTCAAAATTTAGTAC----- 4407
Db 1134 ProGlnIlyLysTyTrpAspGlnValLysLeuAspAspThrSerAspAlaLysThrThrGly 1153
QY 4408 -----AAAGACCACTTTGAGGAATTTGCAACAGACCCAG 4446
Db 1154 SerProIleProSerGlyLysIleThrProLeuThrAspPheIleSerAlaGlyProPro 1173
QY 4447 TTTTACGATGCTTAACCGAATGTACAGCACTATGTATACAGCAAAAAATATTG 4506
Db 1174 TyrPheArgTyTrpLeuGlnIlyTrpGlyLysIleThrPheCysLysGlnIlyGlyAspGly 1193
QY 4507 AAGATGTGCAAGAAAATTAAGTCA----- 4533
Db 1194 GlnLysIleLysGlnIlyCysArgLysAspArgThrGlyHisGlnHisCysSerGlyAsp 1213
QY 4534 -----AATGACCA-----TTGAGTGT 4551
Db 1214 GlyTyrAspCysThrArgTrpAlaAspArgAsnAspLysPheValAspLeuAsnCys 1233
QY 4552 GATACAGATATTAAGAAATCGAGACTGATTAATATATATGAA---AAAAAAA 4608
Db 1234 ---ArgAspCysHisIleGlnCysArgLysTyArgLysTrpIleAspIleLysPheAsp 1252
QY 4609 GAGTGAATTCACAAAGATTAATTTACAGAGTGAACGCAAAAAAGATTCGATAGA 4668
Db 1253 GlnTyHisLysGlnIlyLysIlyLysGlnIlyGlyLysAspLysLeuThrLysAspLys 1272
QY 4669 CAACACATGTGCTAATGCTTACAGACTATACGAAAGTCAACAGATTTACTTGAC 4728
Db 1272 ----- 1272
QY 4729 AGGAAATTTACTGTAGTTGTGTGATTAAGCTGAAAGTCTGTGTGTAACAAAGAAAT 4788
Db 1273 -----SerSerGlyLysAsp----- 1277
QY 4789 ATTCATTTGTAAAAAACAGGCTTACTATGATCCGACAAACATTTGTGGTGCACAAA 4848
Db 1278 -----AsnAsnCysCysLysAspIleGlnIlyHisLysSerAlaIleAlaVal 1292
QY 4849 TTTATGAAATATACAGCAAAATATACATTTTCAGTAAAGATAGTCAAGAGATT 4908
Db 1293 PheLeuLysGlu----- 1296
QY 4909 GTAAGAGGCAAAACACAGTGTATTAAGTGGCAAAACAAAGTCTTATACTAAT 4968
Db 1297 LeuLysHisCysLysAsnGlyGlnThrSer---GlnAsnLysGly---AsnGlnIlyAsp 1314
QY 4969 AACTTGAAGAAATGTACAGATGCTGCTTT---TCTTTCGTCGACTACGATATAG 5024
Db 1315 GlnLeuAsnLysLeuAspPheAspLysIleProGlnIlyThrPheSerProSerThrTyCys 1334
QY 5025 TTTTCATGATGTGAGGCAATATATACAGATCCAGAAAGTAAAGATGAAGAAATGGTTCG 5084
Db 1335 LysAlaCys----- 1337

QY 5085 AAAAGATTCGAGATGCGCCGACGAGAGGGGTACAAATTTGGTCAATATACAAAGA 5144
Db 1338 ---ProVal1TyrglyVal1AsnCySAng1y----- 1346
QY 5145 AAAAAAGAAAAAGAAAAATTAAGATGCGATGCGCACAAATATCTTATGAGTCCC 5204
Db 1347 Asn1ysAr1gGlyAr1gGlyTh1Asn---GlyCysThr1hr----- 1359
QY 5205 GCCTTGTAGTGTCTATGAATATAGTATTTATGATTTAAGAGATTAATTTCTAGGTATGA 5264
Db 1360 -----AsnAsnGlu-ProG1 1364
QY 5265 TAATTGGAAGATGAAAAAGAAAAAGACGAGAAAAATTTGACAAAAATTTAAACAAAA 5324
Db 1364 uAsn1ysGluAsn1sp----- 1369
QY 5325 TGGAAATCATGTTGCAAAAGAAATGATAGTACTACAGA-----AATCCGGTAG 5375
Db 1370 -----LysGlyAla1AserThr1IleSer1Ileu1IleAsn1spGlyse 1384
5376 TACTGCGCGGAAAAATTTTTCGGAACGAAAAATTAAGATGTGTGGAACGCAATGATATG 5435
Db 1384 rThr----- 1385
QY 5436 CGGGTACAAACGTGTAGGATGATGAATATGTGCAAAATGTCAGAAAGTATGAGA 5495
Db 1386 -----AsnGlyAla1ThrAsnGlyThr1GlyThr-----ThrAspGluTh 1399
QY 5496 TCTAAAAAAATGTGTCTGTCTGACCTTCAGATGATGATATCTATG-----GGGAAAA 5549
Db 1399 rLeu1ysGluCys-----SerAsp1ysTy1Ala1PhePhe1ysGlyLeuAr 1414
QY 5550 TCGCGATGAAGTACTGCGCATCATGTTCTTCATGATGTTCCGAATGGGATGAGATTT 5609
Db 1414 g1ysGlnGlu1rThr1CysGln1ys-----LysTy1GlyVal1AsnG1 1428
QY 5610 TTCGAAACATTAAGAAAAAGAAATTTGAGAAATTTGTAGGGCGCTGTAAGATATATCTG 5669
Db 1428 nCys-----AsnLeu1ThrAsnAla1ysAsn1sp---Thry 1439
QY 5670 TGTGTATATGAAGATTAAGAAAAAGAAATGTACAGATGCGGTACCAATATTAATAAT 5729
Db 1439 rPheAsp1ysAsp-----IleVal1PheAsn1GluTh 1449
QY 5730 TATTAGTGTGGAACCAAGATATGAATAAATCAAAATATATGAGATTAAGA 5789
Db 1449 ePheGlnAr1gTTPLeuAr1gTy1Phe-----Val1HisAspTy1Asn1IleLeu1ysH1 1466
5790 CAAAATATATTCGAGCATCTGTGCAAAAGATGCAAGACGCTCGCGAATATTAGA 5849
Db 1466 s1ys1le-----As 1469
QY 5850 CAACAAATTAATAAAATTTGTGAATAATAAGTGAAGATTTGTGAATATATGATGAA 5909
Db 1469 p1roCys1Ile1ysGlu1ysGlu1Asp1ysThr-----Glu1His1ysCys1Ile-- 1485
QY 5910 ACATGTGTCCACACAGCATTAACGTATAGTCAAAATATAGCCCGCATCTTGA 5969
Db 1486 -----Asn1ysCysAsn1le----- 1490
QY 5970 CGATGAACCAAAAGAGTTGAAGAAAGTGTATTTGTCAAGTGCACAGAGTCCACAG 6029
Db 1491 -----LysCysGluCys----- 1494
QY 6030 TGTAGAGAGGAAACACCGCTACACAGGGGTACTGATATCAAAAAGCAGCGCATCGAA 6089
Db 1495 -Val1ty----- 1496
QY 6090 AAAAGAGGAAACAGCGCGCTCTACAACACGCGGAAAAAGTGAATTTCTAACAC 6149
Db 1496 ----- 1496
QY 6150 AGAAATGCGAGCAAAACAGAACCCGACGACACACAAACGAAAAACGAATC 6209

Db 1496 ----- 1496
QY 6210 AACAGCAACACACAGAAATCTGACGTGGGCAAAATGTAAAGCCATCTTTGAAATA 6269
Db 1496 ----- 1496
QY 6270 ACCAGATAGCAGGGGTGAATAGAGGTTGTATCCAAAAACGTATGCAATATCTTA 6329
Db 1497 -----Ly 1497
QY 6330 AAGGGGTATATGTAGTACTTAAGAAAAATGCAATATGATATGATGATGATGATGATG 6389
Db 1497 str1eug1u1Ile1ysGly-----Asn1Glu1rPgly----- 1507
QY 6390 GAGAAAAAATATATGATTAATATATACATATTTAAATATGAAGTGAAGTGAAGTGA 6449
Db 1508 -----Asn1Ile1ys-----LysHisTy1Asn1IleAs 1516
QY 6450 TGCAATGATATTAAGAGAGCGCTTTATTAATGTGCACCAATAGAACTCAATTTTGTG 6509
Db 1516 nSer1Asn1sp1ysGluThr----- 1523
QY 6510 GTTAAATATATATTAATGAATCTGACAGAAAAATGAATTCGAAATGGAACATTC 6569
Db 1523 ----- 1523
QY 6570 ACATGAATTTAAAAATTAATATATATATATATATATATATATATATATATATATAT 6636
Db 1524 -----IleAla1Ty1AsnVal1ysSer1Ty1rPheVal1spGlnGlyLeuPh 1538
QY 6627 TCGAATGATATTTCTAT 6686
Db 1538 eAsp1rAsp1Ty1r1ys1Ala1Gln1ysVal1----- 1549
QY 6687 TCTCATGAATTAATTAAGAAAAAAGCAGATTAAGAAAAAGATGAGATTAAGTAAAT 6746
Db 1550 -----GluAspGlu1ysGluAr1g1ysLy 1557
QY 6747 ATTTGGAGAAAAATTAATAATTTATTTGGGAAGATATATATATATATATATATATAT 6806
Db 1557 s1le1rP-----GlyCysThr1GlyH1 1564
QY 6807 TCTCAGACAGAA---AACGAAAAAGAAAAATTAAGATATATATATATATATATATATAT 6863
Db 1564 s-----Asp1Iuc1ysSer1Glu1ysGlu1ysGlu1Asn1ys-----Asn1Phe1I 1579
QY 6864 GACCAACTGACGCTTCCCTTGAAGAGTTGTAAAAAGCCCAATTTTGGAGATGTT 6923
Db 1579 eThrAsn1Leu1IleSer1Glu1Asn1ys1IeThrSer1CysGln----- 1594
QY 6924 CACAGAAATGGCAGAGAAATTTGTAATAGAGAGAGAACAGTGTTAATAATTTGGAGC 6983
Db 1595 -----Asn1ysHis----- 1597
QY 6984 GGGCTTAAGGAATATAGTATATGTAATGATATATGATATATGATATATGATATATG 7043
Db 1598 -----Asn1ProAsn1Gly1ysThr----- 1603
QY 7044 GGGCTGTATACATATCAAAATTTTATTAAGAAAGTGAAGAACTGAATATGAAGACAAAG 7103
Db 1604 -Ala1Cys----- 1605
QY 7104 AGAAAGTTCAAAAAGATTAAGATGCAAAAGTATTAAGATTTATCTTACTGAAG 7163
Db 1605 ----- 1605
QY 7164 AGACATAGAGAGGCAACATGTGCTCATGAATATTAACATGAATTAAGAAATTAAG 7223
Db 1605 ----- 1605
QY 7224 TGGCAATAGGATTTCTTCTGTATGCAAAACCTTCTTCACTACCAAAACAAACACA 7283

CC Invention. The compositions are for the treatment and prevention of
CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
CC family of genes having homology with conserved regions of DABP and SAMP.
CC The compositions are used for the treatment and prevention of malaria.
CC They are also used in the preparation of vaccines for inducing a
CC protective immune response in a mammal to Plasmodium merozoites
CC (especially Plasmodium falciparum or Plasmodium vivax).

XX Sequence 2710 AA:

Alignment Scores:

Pred. No.:	9.2e-201	Length:	2710
Score:	2972.50	Matches:	865
Percent Similarity:	40.06%	Conservative:	442
Best Local Similarity:	26.51%	Mismatches:	873
Query Match:	15.28%	Indels:	1083
	18	Gaps:	120

US-10-087-013-1 (1-10628) x AAW22482 (1-2710)

25 ATTTAAATGAGGAGATGAGATCATCATTTAGAGGAGATGCTAAAGCCATTATATA 84

Db LeuAlaLysMetGlyProLysGluAlaIleGlyLysAspIleGluAsp----- 22

85 AAGAGAGTCAAGAGAGATGAGATGTTGGAGCTTATGCCAAATATAGACAT 144

Db -----GluteralAlaLysMetPheAspArgIleGlyLysAspValTyrAsp 38

145 CCATCAAAA---TATGCAAGAGAGATGAGATGTTGGAGGATTTGAGCAAGCA 201

Db LysValLysGluGluAlaLysGluArgLysGlyLeuGlnGlyArgLeuSerGlu 58

202 GAATTCGTGGTGGTCCCTTACCGCAGTAAATGATATATATATTCATATCA 261

Db LysPheGluLysAsnGluSerAspProGlnThrProGluAsp-----Pro 73

262 TGTATATAGTCAATGAGAGATGATATATTCAGTATGATGATGATTTGAGCAT 321

Db CysAspLeuAspHisLysTyrHisThrAsnVal-----ThrThrAsnValIleAsn 90

322 CCTTGCATGTTGAGAGAGAGATGATGATGATGATGATGATGATGATGATGAT 378

Db ProCysAlaAspArgSerAspValArgPheSerAspIleTyrGlyGlnCysThrHis 110

379 AATAAATAGTATATATTAAGAGAGAGATGATGATGATGATGATGATGATGATGAT 438

Db AsnArgIleLysAspSerGlnGlnGlyAspAsnLysGlyAlaCysAlaProTyrArg 130

439 CGACATATGTGTATTAAGAGAGATGATGATGATGATGATGATGATGATGATGAT 498

Db LeuHisValCysAspGlnAsnLeuGluGlnIleGluProIleLysIleThrAsnThrHis 150

499 GATTATTTGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 558

Db AsnLeuLeuValAspValCysMetAlaLysPheGluGlnGlnSerIleThrGluAsp 170

559 CATCA---CATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

Db TyrProLysTyrGlnAlaThrTyrGlyAspSerProSerGlnIleCysThrMetLeuAla 190

601 CGAAGTTTTCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660

Db ArgSerPheAlaAspIleGlyAspIleValArgLysArgAspLeuTyrLeuGlyAsnPro 210

661 CATGAC-----AAGTAGAAGAGAGATGATGATGATGATGATGATGATGATGAT 705

Db GlnGluLysIleLysGlnArgGlnGlnLeuGlnAsnAsnLeuLysThrIlePheGlyLysIle 230

706 CATGAT-----GGAATGAGAGATGATGATGATGATGATGATGATGATGATGAT 756

Db TyrGluLysLeuAsnGlyAlaGluAlaArgTyrGlyLysAsnAsp-----ProGlu----- 246

QY 757 GGAATTTATTAATTAAGAGAGATGATGATGATGATGATGATGATGATGATGAT 816

Db -----PheHeuLysLeuArgGluAspTyrThrAlaAsnArgGluThrIleLys 264

817 GCATTAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 876

Db -----AlaIleThrCysAsnAla---TyrGlyAsnThrTyrPhe---HisAlaThrCysAsnArg 282

877 CCATTTATTTCAAAATCCTTAATGCGGCGCAAGAGAGATGATGATGATGATGATGAT 936

Db GlyGluArgThrLysGlyTyrCysArgCysAsnAspAspGlnValProThrTyrPheAsp 302

937 TATGCTCCATATATTTAGCTGTTGCGAGAGATGCGGAGAGATTTGCCAGAGAG 996

Db TyrValProGlnTyrLeuArgTyrPheGluGluTyrPheGluGluAspPheCysArgLys 322

997 AATTTTAATGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1044

Db AsnLysLysIleLysAspValLysArgAsnCysArgGlyLysAspLysGluAspLysAsp 342

1045 CGCTTATATGTTAGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1104

Db Arg---TyrCysSerArgAsnGlyTyrAspCysGluLysThrLysArgAlaIleGlyLys 361

1105 TTGCATTTGATATATAGTACTGATGATGATGATGATGATGATGATGATGATGATGAT 1164

Db LeuArgTyrGlyLysGlnCysIleSerCysLeuTyrAlaCysAsnProTyrValAspTyr 381

1165 TTAGGCAATCAACAGAGATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1224

Db IleAsnAsnGlnLysGlnGlnPheAspLysGlnLysLysLysTyrAspGlnGluLys 401

1225 TCATAT-----TTATCGAAGCATTAACAAATTTGTCAT 1257

Db LysTyrGluLysGlnGlyAlaSerGlySerArgGlnLysArgAspAlaGlyThrThr 421

1258 AATATTAATAGTGA---TATTAACAAATTTATGAGAGAGAGAGAGAGAGAGAGAT 1314

Db ThrThrAsnTyrAspGlyTyrGluLysPheTyrAspIleLeuLysSerGlnTyr 441

1315 GCACTAATGACACTTTTAAATTTACTTAATGAGAGAGAGATGATGATGATGATGAT 1362

Db ArgThrValAspLysPheLeuGlnLysLeuSerAsnGlnIleCysThrLysValLys 461

1363 -----AAGAGGATTAACAGAGAGAGAGAGAGAGATTTACTTAACAGTGTCT 1410

Db AspGlnGluGlyGlyThrIleAspPheLysAsnValAsnSerAspSerThrSerGlyAla 481

1411 -----GATGACAAAGGATATTTATCGTTCAAGATTTGCAAGTGTGT 1455

Db SerGlyThrAsnValGluSerGlnGlyThrPheTyrArgSerLysTyrCysGlnProCys 501

1456 CCCGACGCGCGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1512

Db ProTyrCysGlyValLys-----LysValAsnAsnGlnLysSerSerAsnGlu 517

1513 CGTGAAGCTAATAATAT-----GAAAGTAAAGCTTCATGAGGCT 1554

Db TyrGluLysLysAsnAsnGlyLysCysLysSerGlyLysLeuTyrGluProLysProAsp 537

1555 GTGAAGCTACTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1614

Db LysGluGlyThrThrIleThrIleLeuLysSerGlyLysGlyHisAspAspIleGluGlu 557

1615 AATTTAGAAAATTTTGTAC-----AGCTCA 1641

Db LysLeuAsnLysPheCysAspGluLysAsnGlyAspThrIleAsnSerGlyLysGly 577

1642 ACTAATTAACAAGATAAATAAT-----CAAATGCGAATGC 1680

Db ThrGlyLysSerGlyGlyLysAsnSerGlyArgGlnGluLeuTyrGluGluTyrPheCys 597

1681 TATTTAATAGAGTAAATATATATAGATGTAACTGAGACAAATATGAAATCAATAT 1740

```

Db 598 ----TyrIysGlyIuAspValValIysValIyHisAspGluAspAspGluIuAspTyr 616
OY 1741 GATAATCCTAG-----ATAATA----- 1758
Db 617 GluAsnValIysAsnAlaGlyIleuCysIleuIysAsnGlnIysAsnIysGlu 636
OY 1759 ----- 1773
Db 637 GluIuIyGlyAsnThrSerGluIyGluProAspGluIleGlnIyThrPheAsnProPhe 656
OY 1774 TTGAATATGGGTACATATTTATTAAGGATCTTAAAGGAAATGACAACTTAA 1833
Db 657 PheIyTyrTrpValAlaHisMetLeuIysAspSerIleHisTrpIyLysIysLeuGln 676
OY 1834 ACTTGATA--AATAATACACACACGATTTG--ATTGATGAATCTACAGAAATTCG 1887
Db 677 ArgCysIeuGlnAsnGlyAsnArgIleIyScyGlyAsnAsnIyScyAsnAsnAspCys 696
OY 1888 TTATGTTTACAGATGGGTAAACAAAGAAAGAAAGAAATGAAATGATATTAAGAACTG 1947
Db 697 GluIyShelysArGlyTrpIleThrGlnIyLysAspGluTrpGlyLysIleValGlnHis 716
OY 1948 TTCACAAAAAAGATATACAG----- 1971
Db 717 Phe--LysThrGlnAsnIleIyGlyArgGlyIySerAspAsnThrAlaGluLeuIle 735
OY 1972 ----CAATCGATTAATCTAATATTAATATCTTTTGAAGGTTATTTT----- 2016
Db 736 ProPheAspHisAspTyrValIleuGlnTyrAsnLeuGlnIuIleGluSerGluAspAsn 755
OY 2017 -----TTTAAAGTTATGATTAAGTACAACTGACAAAGAAATGAGAAATG 2058
Db 756 SerGluAspAlaSerGluIuIyLysSerGluAsnSerLeuAspAlaGluIuAla----- 773
OY 2059 AAGAACTTATGGAATAATTAATAAGAAAAAATGAGTTTCCATTTGCAAAATATAT 2118
Db 774 ----GluIuIleuLysHisLeuArgGluIleIleGluSerGluAsnAsn 789
OY 2119 AGGAC-----TATTTAGACATGCAATAGCACTTTGTTA 2154
Db 790 GlnGluAlaSerValIyGlyIyValIThrGluGlnIyLysAsnIleMetAspIyLysLeu 809
OY 2155 GATCACTTAAGAACTGCGACGATATGT----- 2184
Db 810 AsnTyrGluLysAspGluAlaAspIleuAspIeuCysIeuGlnIleHisGluAspGluIuGln 829
OY 2185 ----AAGACATTAATACAAAGACATGTGAACATCCCATTAATGCAACAAACCG 2241
Db 830 LysGluIyGlyAspGluIyAsnGluCysIleGluGlnIyGluAsnPheArgTyrAsnPro 849
OY 2242 TGTGTTAAACCTCGGAGCGACGAA---CCGACTAAATAATTAAGAAATGACCA 2298
Db 850 CysSerGlyIuSerGlyAsnIyAsnIyGlyTrpValLeuAlaAsnIyValAlaTyrGln 869
OY 2299 TACTTAAAGAGATGCATACGAGAGACGAAATCGTGTCTCATTAATTTGAAGA 2358
Db 870 MetHisIyLysAlaIyThrGlnIleuAlaSerArgIyValArgSerAlaLeuArgIy 889
OY 2359 AAGGACACGAGATTAATTAAGTGGGGTAGAGAAAGACTTCAGACAAATTA 2418
Db 890 AspIleSerLeuAlaGlnPheIyLysAsnGlyArgAsnGlySerThrLeuIyScyGlnIle 909
OY 2419 TGTAGATTAATGATTAACATCTTAATCTGTTTTCATTAAGGCAACGATAT 2478
Db 910 CysLysIleAsnGluAsnTyrSerAsnAspSerArgIyAsnSerGlyIyProCysThr 929
OY 2479 GGCACAGCAGCAGTGTGATTAACAAACAGATTTGCTAGAGAACTGAGGAACTG 2538
Db 930 GlyLys---AspGlyAspHisGlyIyValAlaArgMetArgIleGlyThrGluTrpSerAsn 948
OY 2539 GATCGGAACACATGCGTAAGATACAGAAATGTTATATGCTCGTGAAGACGACAT 2598

```

```

Db 949 IleGluIyLysGlnIleThrSerTyrLysAsnValPheLeuProProArgArgIuHis 968
OY 2599 ATATGATACATCCATTTGGACATTTTACAAACGATGATACCCACTTAATGTAATAT 2658
Db 969 MetCysThrSerAsnLeuGluAsnLeuAspVal-----GlySerVal 982
OY 2659 GTT---GATGATTAGTATATATTCCTTTTGGGGGAGCTTCTATACGCAAAATAT 2715
Db 983 ThrIyAsnAspLysAlaSerHisSerLeuLeuGlyAspValGlnIleuAlaIyThr 1002
OY 2716 GAACAAACAGATTAATACGAATGATAAAGAAAGAAATTAACCTAAAGGCCCAAGAA 2775
Db 1003 AspAlaIleGluIleIleLysArgTyrLysAspGlnAsnIle-----Gln 1018
OY 2776 GTAACGACCA-----AACAACGACACACTATCTGCGACTATACGTTACAGT 2826
Db 1019 LeuThrAspProIleGlnGlnIyAspGlnIuAlaMetCysArgAlaValArgTyrSer 1038
OY 2827 TTGCGATATAGTGAATTAATTCGAGAGAGATCTCTGGGAAGAAAGGTCATAG 2886
Db 1039 PheAlaAspLeuGlyAspIleIleArgIyArgAspMetTrpAspGluAspLysSer 1058
OY 2887 GTAACGTCGAAGACATTTGAAAACGTGTTTGTAAATATACATAGCTCAACAGC 2946
Db 1059 ThrAspMetGluThrArgLeuIleThrValPheLysAsnIleLysGluIyHisAspIy 1078
OY 2947 ----AAGGAATGATTAATTAATGATGAT-----GCCCAATATTTAAATG 2994
Db 1079 IleIyAspAsnProLysTyrThrGlyAspGluSerLysIyProAlaTyrLysIyLeu 1098
OY 2995 AGGAAAATTTGGTGGAGAGCTAATAGACCAAGTAAAGTGAAGCAATGATGATTA 3054
Db 1099 ArgAlaAspTrpTrpGluAlaAsnArgHisGlnValIleTrpArgAlaMetLysCys----- 1116
OY 3055 AATATTTGAAGGATTAATCGGACACCAATCAACACAAATAGTTATTCGGATTAAT 3114
Db 1117 -----AlaTrpIyGlyIleIleCys----- 1123
OY 3115 GATCATACACATGATGATTAATCCACAAATAATTAATGATGATGACGATGGCA 3174
Db 1124 ProGlyMetProValAspAspTyrIleProGlnIleuArgIleuArgTrpMetThrGluTrpAla 1143
OY 3175 GATGTGCTCGACAGCTCAGAAAAGAGATGATTAAGTTGAAGAGAGAGTGAAG 3234
Db 1144 GluTrpTyrCysLysAlaGlnSerGlnIuIyTrpAspIyLysIleCysAlaAsp 1163
OY 3235 TGTAAAGATAGGATTAATGCTCAGGCTGACGAAAGAGAGTGTACAGTGTGACAG 3294
Db 1164 CysMetSerLysGlyAspGlyLys---CysThr---GlnIyAspValAspCysGlyLys 1181
OY 3295 TGCACAGAGCTTGTAAATTAATGATTAATGATTAATGATTAATGATTAATGAT 3354
Db 1182 CysIyAlaIaIyAspLysAspLysTyrGlnIuIleGlnIyLysTrpAsnGluIuIleTrpArg 1201
OY 3355 ATATATCAGATTAATTAATTAAGATTAATTAATTAATTAATTAATTAATTAAT 3414
Db 1202 LysIleSerAspLysTyrAsnLeuIyLeuIleAlaIyLysThrSerThrAsnPro 1221
OY 3415 GGTATTAAGCTTCCAGTATGCCCAAAATCATATAGAC-----AGCAATGTTAT 3465
Db 1222 Gly-----ArgThrValLeuGlyAspAspAspProAspTyrGlnIleMetVal 1237
OY 3466 GAATTTTTCGCAATTAATACCAAAATGCTGGCA-----AGTAATTA 3513
Db 1238 AspPheLeuThrProIleHisLysAlaSerIleAlaIleArgValLeuValLysArgAla 1257
OY 3514 AGTGTACTAGTGTAAAGGCTGATGATTAATGATTAATGATTAATGATTAATGAT 3573
Db 1258 AlaCysIleProThrGluIleAlaAlaIleAlaProIleThrProTyrSerThrAlaAla 1277
OY 3574 GCATATCTCATGATTAACGAGAAATTTGATGATGATGATGATGATGATGATGAT 3633
Db 1278 GlyTyrIleHisGlnIuIleGlyTyrGlyIyCysGlnIuIleIleThrGlnPheCysGlu 1297

```


Db 1704 ----- 1704
QY 5803 GAGCATCCTGGCAAAAGATCGAGAGAGCGCTCGCAATATTAGCAAAATTTAAA 5862
Db 1704 ----- 1704
QY 5863 AAAATTTGTGAAAAAGTGGAGATTGTGAATATAGTATGAAAGATGTCCACA 5922
Db 1705 ----- 1706
QY 5923 CAGGATTAACTGTGTATAGCAAAATATGCCCGCATCATATTAGCATGAACCAAA 5982
Db 1706 ----- 1706
QY 5983 GAAGTTGAAGAAAAGTGAATTTGCAAGTCCACAGAGTCCACAGCGTGTACGAGGGA 6042
Db 1707 ----- 1717
QY 6043 ACACCGTCACACGGGTATCATGTATCAAAAAGCAGCGCATGAAAAAGAGCGAAA 6102
1717 ----- 1717
QY 6103 ACAGCGCGCTACAAAAGCAGAAAAAGTGAATTTACACACAGAAATCCAGCA 6162
Db 1718 ----- 1725
QY 6163 CAACAACGAAACCGAGCAGACGACAAACAGAAAAAGAACATCAACAGCAACAGA 6222
Db 1726 ----- 1730
QY 6223 ACAGAACTGACGTGGGCAACATGTAAAGCCATCTTTCGAATTAACCAAGATACAG 6282
Db 1731 ----- 1747
QY 6283 GGTGAATAGAGGGTGTGTAATCCAAAACGTATGACAAATAT ----- CCTAAATGGGGT 6336
Db 1748 ----- 1766
QY 6337 TGTATTT ----- GTAGTAAGTCTAAAGAAAT 6363
Db 1767 ----- 1786
QY 6364 GAAATGCGATATGTATGCGCTTCAGCAAAATTAATTAATTAATTAATTAATTAAT 6423
Db 1787 ----- 1805
QY 6424 TTAATTTATGAACT ----- GAAATTAAGCGTGAC 6453
1806 ----- 1824
QY 6454 AATGATATTAAGAGCGCTTTTATTAATGTGAGCAATATGAACATTTTGTGCTTA 6513
Db 1825 ----- 1844
QY 6514 AAATATTAATTAATGAA ----- AATCTGACAGCAAAATGAA ----- 6549
Db 1845 ----- 1864
QY 6550 ----- TGCMAAATGGAACA 6564
Db 1865 ----- 1884
QY 6565 ATTCAGATGAATTAAGAAATATGTATATATATATATATATATATATATATATATATAT 6624
Db 1885 ----- 1904
QY 6625 TTTGGAAGTATATTTCAATGATTAATAATTAATTAATTAATTAATTAATTAATTAAT 6684
Db 1905 ----- 1924
QY 6685 ATTCATCAATTAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 6744

Db 1925 ----- 1944
QY 6745 ATA ----- 6747
Db 1945 ----- 1964
QY 6748 ----- TTTGGAGGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6804
Db 1965 ----- 1984
QY 6805 ----- CATCTCACAGCAAGAAACGAAAAAGAAATTTGA 6840
Db 1985 ----- 2004
QY 6841 GATTAATTAAC ----- 6849
Db 2005 ----- 2024
QY 6850 ----- CAGTACAAAT ----- GACATGACCAACATG ----- 6873
Db 2025 ----- 2044
QY 6874 ----- ACCCTTCCTTGAGAGTGTGTAATAAGGCC 6906
Db 2045 ----- 2064
QY 6907 CAATTTTGAGATGAGTTCACAGATGCGCAGACAAATTTTGAATTAAGAGAGAACAG 6966
Db 2065 ----- 2084
QY 6967 TTTGTAATAATGAGAGCGGCGGTGAAGAAATATGAG ----- 7002
Db 2085 ----- 2104
QY 7003 ----- TGTAAAT ----- GGTAGTAAT 7017
Db 2105 ----- 2124
QY 7018 GACGCTAAGACAAAGAA ----- TGTGAGAGCGGTGTGTAACATATCA 7062
Db 2125 ----- 2144
QY 7063 AATTTTAAAGAGTGAAGAACTGAATATGAAGAAAGAAAGAAAGTTCAAAAAGAT 7122
Db 2145 ----- 2164
QY 7123 AAAGAT ----- 7128
Db 2165 ----- 2184
QY 7129 ----- GCGAAAAAGTATAGGATTAAT ----- CCTTCTACTGAAGAGACATA 7170
Db 2185 ----- 2204
QY 7211 GAGAAAGCAACATGCTCATGATATTAACATGAATTAATAAGAAATTAATGTCGAAT 7230
Db 2205 ----- 2219
QY 7231 AAGGATGCTCTGTATGACAAACCTTCTTCACAACTACCAAAACACACAAACATCA 7290
Db 2220 ----- 2234
QY 7291 CAATCATCGATGCTATGATATGCCAAGATCCGTGATTAATGTCCTGAAGATTTAAC 7350
Db 2235 ----- 2248
QY 7351 AAGGTGAGTGTCTGACATTTCAAAAAAGGATCTATGATTAACAAAAAATTAAT 7410
Db 2236 ----- 2248
QY 7411 GAACTTAATTAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 7470
Db 2249 ----- 2256
QY 7470 ----- AATATTAATTAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 7530

QY	7471	GAATATATATGATATACATTCGTAAGGAAAGAAATTTATATACATATATGATGCTACAAAGAA	7530
Db	2257	GIunsnllyValaspserThrVal-----LeuGIuMeArlyValSer	2270
QY	7531	AAGGAAGTAAATAGTGGACTAAATTAATTCCTGGCATCTAAGAAACCTTATGCA	7590
Db	2271	AlaspserIysSerGIyPheCnclYAsp-----	2280
QY	7591	CCTGATTAATATATAGGAGAAAGAAACCTTGTAAGAT-----	7629
Db	2281	-----GlyLeuGIuasnAlaCysArvgIyAlaGlyIlePheGluGlyIle	2295
QY	7630	AGAGAGAAATCGTTTAAAGCTAGATTTATGAGTAATTCAGAAATTCACAAAGTTC	7689
Db	2296	ArgIysAsp-----GluTrpIysCys--ArgAsnValCysGly	2307
QY	7690	TATACAGAGAGAAAAAGAGATATGTACTCTCCACAGACAGAAACATATGTCTTAAGCAT	7749
Db	2308	TyrVal-----ValCysIysPro-----	2313
Db	7750	TTAGATGAATTAATTAATGAAAGCTTAAGATAGATTAATTCCTTAATAATGGTTCG	7809
Db	2313	-----	2313
QY	7810	CGAAGCTCAGAAATGAGAGAAATAGACATATTAATAAACTTCACACTCAGAGACGGTGC	7869
Db	2314	-----GluAsnValAsnGIyGlu-----	2319
QY	7870	GCATGAAATCCCATATCTGTGATACTATAGAAATATAGTTTCGTGATCTGGGTGCATAGT	7929
Db	2320	-----Ala	2320
QY	7930	AGAGAGACAGATATGTACGAATGTGGTACTTACTCCCGGTGAATTAATTAAT	7989
Db	2321	LyselIyLysHisIleIleGlnIleArgAlaIleuValIys-----ArgTrpVal	2336
QY	7990	AAGGTTTGAATACATATATGTGAAATATGAGAAATTAATAAGGTAGAAATTAATAC	8049
Db	2337	GIuIyTrpPheGluAsnIyTrpAsnIyIleYsHisIyIleSerHisArgIleIysAsn	2356
QY	8050	AAGCAT-----GTACAAAGCTTTCGTTCCGTTCGTGGGAGCTAATAGAAA	8097
Db	2357	GIyIuIleSerProCysIleIysAsnCysValGIuIyStrpValAspGlnIlyArgIys	2376
QY	8098	GATATTTGGAAGCAATGACGTGCAAGACACCAAGAGATGCAAACTTTTGAAGAAAGA	8157
Db	2377	GIu-----TrpIysGIuIleThrGIuArgPheIysAsp-----	2387
Db	8158	AGATGATGATGATTTGAAGCATTAATTAACAAGATTAAGTGGACATTAAGACAT	8217
Db	2388	-----GluTyrIyLysAsnAspIysSerIysAsp	2396
QY	8218	CCACCTGTTGATGATTAATATATACCTCAACGGTTTCGATGATGATGAATGTGCTGAATAT	8277
Db	2397	AspAsnValArgSerPheIeuGIuThrIleuIleProGlnIleThrAspAlaAsn-----	2414
QY	8278	TATGTGTAACACTGATGAGAAAGATTTGAAAAATTTAAAAATCTCTGATCAGCTGTAA	8337
Db	2415	---AlaIysAsnIyValIleIyIleuSerIySPheGIyAsnSerCysGly--CysSer	2432
QY	8338	ACATCTACAGATGACGAAGATGATTAATGATGAATTAAGTGGACAGCTGTAAACGAGA	8397
Db	2433	AlaSer-----AlaAsnGIuGlnAsnIyAsn-----	2441
QY	8398	TGTCAACAATATATAAATTTTGTTCTTAAATGAGAAATCTCTATTCGATATACATCAAT	8457
Db	2442	---GlyGIuIyTrpIyAspAlaIle-----AspCysMeIeuIys	2453
QY	8458	AATATCAAGAAATGTATGAAACAACAATATATACAAAAATGCTACTTAATGATCATGTT	8517
Db	2454	LysIeuIyAsp-----LysIleGIyGluCysGIuIyIys	2465

[illegible]

	RESULT 6
AA77904	
ID	AA77904 standard; Protein; 2710 AA.
XX	
AC	AA77904;
XX	
DT	13-JUN-2000 (first entry)
XX	
DE	P. falciparum Proj3 binding domain polypeptide.
XX	
KW	DBI gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KM	DABP; Sialic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
XX	protozoacide; Proj3.
XX	
OS	Plasmodium falciparum.
PN	
XN	US593827-A.
XX	
PD	30-NOV-1999.
XX	
PF	07-JUN-1995; 95US-0487826.
XX	
PR	10-SEP-1993; 93US-0119677.
XX	
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.
XX	
PI	Salm KL, Chittins C, Peterson DS, Su X, Wellens TE, Miller LH;
XA	
DR	WPI; 2000-194198/17.
DR	N-PSTDB; AAZ98286.
XX	
PT	Isolated protein binding domains from Plasmodium vivax and Plasmodium falciparum erythrocyte binding proteins useful for vaccinating against malaria -
PS	
XX	Disclosure; Columns 79-92; 93pp; English.
CC	
CC	The invention relates to ebl-1 polypeptides that are encoded by the DBL (Duffy-binding like) gene family. The ebl-1 proteins are substantially identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid Binding Protein (SABP), which are soluble proteins that appear in the culture supernatant after erythrocytes infected with malaria release merozoites. Immunohemical studies indicate that DABP and SABP are the respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be used to vaccinate against malaria, especially caused by P. falciparum. Immunization with the polypeptide provides effective protection against malaria. The present sequence represents the Proj3 binding domain polypeptide.
CC	
CC	
XX	Sequence 2710 AA;
SO	
	Alignment Scores:

Pred. No.: 9.2e-201 Length: 2710
 Score: 2972.50 Matches: 865
 Percent Similarity: 40.06% Conservative: 442
 Best Local Similarity: 26.51% Mismatches: 873
 Query Match: 15.28% Indels: 1093
 DB: 21 Gaps: 120

US-10-087-013-1 (1-10628) x AAY77904 (1-2710)

```

OY 25 ATTTAAATGAGGATGACATCATTCATTTAGAGGAGATGCTAAAGCCCTATTATA 84
   :: |||||
Db 6 LeuAlaLysMetGlyProLysGluAlaIleGlyCysAspIleGluAsp----- 22
OY 85 AAAGAAGTCACAAAGATGCAAGAAATGTTTGAGCGATTATGCCAAATATATAGACAT 144
   ::|||
Db 23 -----GusSerAlaLysHisMetPheAspArgIleGlyLysAspValTyrAsp 38
OY 145 CCATCAAAA--TATGCAAAAAGAACATGTGATTCGTGAAAGGATTTGAGCAAGCA 201
   ||| |||||
Db 39 LysValLysGluGluAlaLysGluArgGlyLysGlyLeuGluGlnIleArgLeuSerGluAla 58
202 GAATTCGCGTGGTCCCTCTACGCCAGTAATATACATATTTATTATTCATATCCA 261
   ::||| ||| |||
Db 59 LysPheGluLysAsnGluSerAspProGlnThrProGluAsp-----Pro 73
OY 262 TGTATTTAGATCATAGAACATCTACTATTTACGGTATGATGATGTGATTTGAGACAT 321
   ||||| |||||
Db 74 CysAspLeuAspHisLysLysTyrHisThrAsnVal-----ThrHisAsnValIleAsn 90
OY 322 CCTTGCATGTAGAGAACAAACCGATTGTATGATGAGATGAACATCTGAAATCT 378
   ||||| ||| |||||
Db 91 ProCysAlaAspArgSerAspValAlaArgPheSerAspIleGlyLysGlnCysThrHis 110
OY 379 AATTAATATCGTATATTAATAAGAAAAAATGATGCTATAGCTCTGTGCGCCACTAGACA 438
   ||||| |||||
Db 111 AsnArgIleLysAspSerGlnGlnIleGlyAspAsnLysGlyAlaCysAlaProTyrArgArg 130
OY 439 CGCATATGTGTGATTAATAAACTGGAAAGCTCTAATGATATTAATATCCAAATATTCAT 498
   ||||| |||||
Db 131 LeuHisValLysAspGlnAsnLeuGluGlnIleGluProIleLysIleThrAsnThrHis 150
OY 499 GATTTATGGGAATTTACTAGTTACAGCAAAATACAGAGGATGATTCATTTGATTAAT 558
   ||||| |||||
Db 151 AsnLeuLeuValAspValLysMetIleAlaLysPheGluGlnIleGlnSerIleThrGlnAsp 170
OY 559 CATCCA---CATTAAGGAACCT-----TCAGACGCTTGATCTCTTGCA 600
   ||||| |||||
Db 171 TyrProLysTyrGlnAlaThrTyrGlyAspSerProSerGlnIleCysThrMetLeuAla 190
601 CCAAGTTTGCAGATATAGTATTTGTAAGAGAAATGATGATTTTAAACCAATGTC 660
   ||||| |||||
Db 191 ArgSerPheAlaAspIleGlyAspIleValAlaArgGlyArgAspLeuTyrLeuGlnLysPro 210
OY 661 CATGAC-----AAAGTAGAAACGGGCTCCGAGAGGATTTTCAAGAAATA 705
   ::||| |||||
Db 211 GlnGluIleLysGlnArgGlnIleGlnLysAsnAsnLeuLysThrIlePheGlyLysIle 230
OY 706 CATGAT-----CGAATGGAACATGAAGTAAATATGATTTACATCTCGATGATCT 756
   ||||| |||||
Db 221 TyrGluLysLeuAsnGlyAlaGluAlaArgTyrGlyAsnAsp-----ProLys 246
OY 757 GGAATTTATTAATTAAGAACACATGCTGATGATGATGATGATGATGATGATGATG 816
   ::||| |||||
Db 247 -----PhePheLysLeuArgGlnAspTyrPyrThrAlaAsnArgGluThrValTyrLys 264
OY 817 GGTATACATGTGATGATCATATTAATCTGATTTATTTATGCAATCAGAAATATACA 876
   ||||| |||||
Db 265 AlaIleThrCysAsnAla---TyrGlyAsnThrTyrPhe---HisAlaThrCysAsnArg 282
OY 877 CCATATTTTCAATCTCAATATGCGGCATTAACAGAGAAAGTCTCTACCAATTTAGAT 936
   ::||| |||||
Db 283 GlyLysArgThrLysGlyTyrCysArgCysAsnAspAspGlnValProThrTyrPheAsp 302
  
```

```

OY 937 TATGCTCCATATATTTAGCTTGTCGACGAATGGGAGAAAGTTCGCCGAAAGCA 996
   ||||| |||||
Db 303 TyrValProGlnIleuThrLeuThrPheGluGluTyrPalaGluAspPheCysArgLysLys 322
OY 997 AATATTAATTTGAAAAAGGCTCAAGACCTCTGCTCT-----AATGACAAAGAA 1044
   ||||| |||||
Db 323 AsnLysLysIleLysAspValLysArgAsnCysArgGlyLysAspLysGluAspLysAsp 342
OY 1045 CGCTTATATTTAGTCATATATGACATGATGTTGTACGACAACTATTTGCAAAAGCTATT 1104
   ||| |||||
Db 343 Arg---TyrCysSerArgAsnGlyTyrAspCysGluLysThrLysArgAlaIleGlyLys 361
OY 1105 TTGCATTTGATTAATTAAGTACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1164
   ||| |||||
Db 362 LeuArgTyrGlyLysGlnCysIleSerCysLeuTyrAlaCysAsnProTyrValAspTyr 381
OY 1165 TTAGGGAATCAACAGAACATTTAAAAAAGAAAAAAGAAATATGAAAGAAATACA 1224
   ::||| |||||
Db 382 IleAsnAsnGlnLysGluGlnIlePheAspLysGlnLysLysTyrAspGluIleLys 401
OY 1225 TCATAT-----TTATGCAACGATTAACAAATTTGTCAT 1257
   ||| |||||
Db 402 LysTyrGluAsnGlyAlaSerGlyLysSerArgGlnLysArgAspAlaGlyTyrThr 421
OY 1258 AATATTAATATAGGAA---TATATTAACAAATTTATGAAAACTTAAGAAACGCAATAT 1314
   ||| |||||
Db 422 ThrThrAsnTyrAspGlyTyrGluLysLysPheTyrAspGluLeuAsnLysSerIleTyr 441
OY 1315 GCACTAATGACACTTTTAAATTTACTAAATGAAGAAAGTATGT----- 1362
   ||| |||||
Db 442 ArgThrValAspLysPheLeuGluLysLeuSerAsnGluGlnIleCysThrLysValLys 461
OY 1363 -----AAAGGAGATTTACCGAGAGAAAGATATTTACTTTTACTAACAGTCT----- 1410
   ::||| |||||
Db 462 AspGluGluGlyGlyThrIleAspPheLysAsnValAsnSerAspSerThrSerGlyAla 481
OY 1411 -----GATGACAAAGGATTAATTTATGCTTGAATATTTGCCAAGTGT 1455
   ||||| |||||
Db 482 SerGlyThrAsnValGluSerGlnGlyThrPheTyrArgSerLysTyrCysGlnProCys 501
OY 1456 CCGGACCTCGGGGCTCAATGTCATGATGATTAATAATACACACAC---AATCAGATTAATGAT 1512
   ||| |||||
Db 502 ProTyrCysGlyValLys-----LysValAsnAsnGlyLysSerSerAsnGlu 517
OY 1513 CGTGAACCTGTAATAAT-----GAAGCTATTAACCCCATGGGGT 1554
   ||| |||||
Db 518 TrpGluGluLysAsnAsnGlyLysCysLysSerGlyLysLeuTyrGluProLysProAsp 537
OY 1555 GTGAAGCTACTAATATCATCGTCCCTTATATAGTATGATGAACAGGTGATATACAA 1614
   ::||| |||||
Db 538 LysGluGluTyrThrThrIleThrIleLeuLysSerGlyLysGlyHisAspAspIleGluGln 557
OY 1615 AATATGAAGAAATTTTGTGAC-----AGCTCA 1641
   ||||| |||||
Db 558 LysLeuAsnLysPheCysAspGluLysAsnGlyAspThrIleAsnSerGlyLysSerGly 577
OY 1642 ACTAATTAACAAGATTAATAATAT-----CAAAATGGGAATGC 1680
   ||||| |||||
Db 578 ThrGlyLysSerGlyLysGlyAsnSerGlyArgGlnIleuLeuTyrGluGluThrLysCys 597
OY 1681 TATATTAAGGATGAATAATTAATATATGATGTAACCTGCAACAAATATCAATTAAT 1740
   ||||| |||||
Db 598 ---TyrLysGlyLysAspValIleValLysValGlyHisAspGlnAspAspGluGluAspTyr 616
OY 1741 GATAATCTTAAG-----ATATA 1758
   ::||| |||||
Db 617 GluAsnValLysAsnAlaGlyLysLeuCysIleLeuLysAsnGlnLysLysAsnLysGlu 636
OY 1759 -----TCATTTCAATATTT 1773
   ::||| |||||
Db 637 GluGlyLysAsnThrSerGluLysGluProAspGluIleGlnLysThrPheAsnProPhe 656
OY 1774 TTGAAATTATGGGTTACATATTTTATTAAGGATTAATTAAGTGAATGACAAACTTAA 1833
  
```


OY 3796 GATATACCTTAAGAAAAACGATGGAAGAAACAGTACAGATGTCATCCAAAAAGAT 3855
Db 1354 LysIleLeuGlnGlyLysAsnGlyArgThrValGlyGlyCysAsnProLysGlnSer 1373
OY 3856 AGTATGATATCCCGATGGCAATGC---GGAATATTAATTTAAGTGAAGACCCGCT 3912
Db 1374 -----TyrProAspTyrPaspCysLysAsnAsnIleAspIleSerHisAsp---Gly 1389
OY 3913 GTGTGTATGCCCCCTAGACAGAAAGTATGCTACATTTCTGGAAATGATATGAA 3972
Db 1390 AlaCysMetProPheArgGlnGlnLysLeuCysLeuTyrTyrIleHisGlnSerGln 1409
OY 3973 ATAAAAAATTACATCACAAGTAAATTTAAAGACCTTTCAAAATCGACAGCA 4032
Db 1410 ThrGlnAsnIleLysThrAspAsnLeuLysAspIleHisLeuThrIleAlaAla 1429
OY 4033 GAAACATTTCTTCAGTGTATTTATTAAGATAG---GATGGTGAAGAAATGAACCTC 4089
Db 1430 GlnThrPheLeuSerTyrGlnTyrTyrLysSerLysAsnAspSerGlnAlaLysIleLeu 1449
OY 4090 GATAAAGATTTAAAGAGGCAAAATTCCTCCGCAATTTTGAGATCCATGCTACACA 4149
Db 1450 AspArg-----GlyLeuIleProSerGlnPheLeuArgSerMetCysTyrThr 1465
OY 4150 TTTGGAGATTATGAGATTTTATTTATTTGACACAGATATATCAAAAGTCATGGTGAGGA 4209
Db 1466 PheGlyAspTyrArgAspIleCysLeuAsnThrAspIleSerLysGlnAsnAspVal 1485
OY 4210 AGTAACTTAAGACAGATAGATTCCTTTCAAAAAATGGTACCAAAATCTCTAAT 4269
Db 1486 AlaLysAlaLysAspLysIleGlyLysPhePheSerLysAspLysSerProSer 1505
OY 4270 GGA AAAACACGCCAAGATGCTGACAGACATAGTCATGATATGGAAGCTATGCTA 4329
Db 1506 GlyLeuSerArgGlnGlnTyrPheLysThrAsnGlyProGlnIleTyrPheGlnLeu 1525
OY 4330 TGTGCACTAGTAAATTTGGGGCAAAAAAGATGAT-----TTTACCGAAAATC 4380
Db 1526 CysAlaLeuThrLysTyrValThrAspThrAspAsnLysArgLysIleLysAsnAspTyr 1545
OY 4381 GGTTCACACACGCAATTAATTTAGTGACAAA---AGCACCACTTGGAGAAATTTGGCA 4437
Db 1546 SerTyrAspLysValAsnGlnSerGlnAsnGlyAsnProSerLeuGlnIlePheAlaAla 1565
OY 4438 CGACCCCACTTTTACGATGCTACCGCAATGCTACGACGACTATTTGCTATACAGACA 4497
Db 1566 LysProGlnPheLeuArgTyrPheTyrIleGlyTyrPheGlnGlnPheCysAlaGlnArg 1585
OY 4498 AATATTTGAGAGATGCGACGAGAAAAATGTAAGTCA---AATGACCAATTTGAAGTGTAT 4554
Db 1586 LysLysGlnAsnIleLeuLysAspAlaCysAsnGlnIleAsnSerThrGlnIleCysAsn 1605
OY 4555 ACA-----GAATTTTAAAGAAATGCGAGACTACGCTTAATTAATATGAAA---AAA 4602
Db 1606 AspAlaLysHisArgCysAsnGlnAlaCysArgAlaTyrGlnGlnTyrValGlnAsnLys 1625
OY 4603 AAAAAGATGGATCCACAGATATAATTTACAGAGATGAACGGCAGACAAAAAGATTC 4662
Db 1626 LysLysGlnPheSerGlyGlnThrAsnAsnPheValLeuLysAlaAsnValGlnProGln 1645
OY 4663 GATACACACACATTTGCTATGCTATGCTACAGACTACTGGAACGAATGCAACAGATTAC 4722
Db 1646 AspProGlnTyrLysGlyTyrGlnTyrLysAsp----- 1656
OY 4723 TTGAACAGAAATTTAGTGTAGTTGTGTATGAACCTGGAAGTCCCTGTGTATACA 4782
Db 1657 -----GlyValGlnPro-----IleGln 1662
OY 4783 AGAATATATCAATTTGTAGAAAAACAGCCTTACTATAGTCCGCAAAACATTTGGGGTGC 4842
Db 1663 GlyAsnGlnTyrLeuLeuGlnLys-----CysAspAsnAsnLys---CysSerCys 1678

OY 4843 ACAAAATTTATGAAATGACGACAAATATCTAACATTTGAGTAAGATAGTGCAAA 4902
Db 1678 ----- 1678
OY 4903 GGATAGTAAAGAGGCAAAACAGGTGCTATTAGTGGCAAAACAAAGTCTTAATAC 4962
Db 1678 ----- 1678
OY 4963 TACAAATTACTGAAGAAATTTAGCTAGATGTCTTTTCCCTTCGTGACAGTATA 5022
Db 1678 ----- 1678
OY 5023 TGTTCATGATGGATGGCAATTATACAGATCCAGAACTTAAGATGAAGAAATGGTTG 5082
Db 1679 -----MetAspLysAsnValLeuSerValSerProLysGlnLys----- 1691
OY 5083 CGAAAAAGATGATGGAAGTGGCGCAACGAAAGGGTCAATTTGGTGCATACTACAAA 5142
Db 1692 -----ProPheGlyLysTyr----- 1696
OY 5143 GAAAAAAGAAAAAGAAAAATTAACGTCGATGGCACAAATATTTATGAGGTC 5202
Db 1697 -----AlaHisLysTyrPro----- 1701
OY 5203 CCGCCTTGATGCTATGAATATAGTTTATGATTTAAAGATATATTTCTAGTATT 5262
Db 1701 ----- 1701
OY 5263 GATTAATTTGAAAGATGAAGAAAAACAAAGACGAGAAATTTGAAGAAATATTAACAA 5322
Db 1701 ----- 1701
OY 5323 AATGAACATAGTTGGCAAGAAAGATAGTACTACAGAAATCCCGTAGTACTCG 5382
Db 1701 ----- 1701
OY 5383 CGAAATTTTCTGGAACGAATTAAGAAATGTGTGTGGAACGCAATGATGGGGTAC 5442
Db 1701 ----- 1701
OY 5443 AAACGTGTAGGATGATGGAATATGTGAATATGTCAAGAAAGTGAAGATCTAAA 5502
Db 1701 ----- 1701
OY 5503 AAATGTGTTCTGTACTTCATGATGATATATCTATGGGAAAAATCCGATGAAGT 5562
Db 1701 ----- 1701
OY 5563 ACTGCGTATCAGTTTCTGATGTTGCCGAATGGGCGTGAAGATTTTGCAAACATAAA 5622
Db 1701 ----- 1701
OY 5623 GAAAGGAATTTGGAATTTGTAAGGGCGTATGATGATTAATCTGTGTGTAATGAA 5682
Db 1701 ----- 1701
OY 5683 GATTAAGAAAGAAATGTACAGATGCTGTACACAAATATTAATTAATTAAGTGG 5742
Db 1702 -----GluLysCys----- 1704
OY 5743 AAACACAGTATGAAAAACAATCAAAAAATATATGTAGATAAAGCAAAATATATTC 5802
Db 1704 ----- 1704
OY 5803 GAGCATCTGTGGCAAAAGATGACAGAGACGCTCGGAATATTTAGACAAACATTAATA 5862
Db 1704 ----- 1704
OY 5863 AAAATTTGTAATAATAAGTGAATTTGATATATAGTATGAAGATGTGCACA 5922
Db 1705 -----AspCys----- 1706
OY 5923 CAGCATTAAGTATGATGTAATATGCAAAATATGCCCCGCAATCAATTAGACGATGAACCAAA 5982

Db 1706 ----- 1706
QY 5983 GAAGTTGAAGAAAGTAAATGTCAGAGTCCAGAGTCCAGCTGTACGAGGAA 6042
Db 1707 ---TyrGlnGlyLys-----HisValProSerIleProPro----- 1717
QY 6043 ACACCGTCACACGCGGTATCAGATATCAAAAGCGAGCATCGAAAAAGAGCGAAA 6102
Db 1717 ----- 1717
QY 6103 ACAGCGCCGCTCAAAAACAGCCGAAAAAGTGAATAATCTAACACAGAAATGCCAGCA 6162
Db 1718 ---ProProProProValGlnProGln----- 1725
QY 6163 CAACACAGAACCCGACGAGCAGCACAACAAACAGAAACGACATCAACAGCACAACCA 6222
Db 1726 ----- 1730
QY 6223 ACAGAAATCTGACGTGGGACAAATGTAAGCCATCTTCGAATAAACACAGATACAGAG 6282
Db 1731 ValThrValAlaSerValLysSerIleValLysThrLeuPheLys-----AspThrAsn 1747
QY 6283 GGTGGAATAGAGGTTGTATCCAAAACGTATGACAAATAT-----CCTAAATGGGGT 6336
Db 1748 AsnPheSerAspAlaCysGlyLeuLys---TyrGlyLysThrAlaProSerSerIlePheLys 1766
QY 6337 TGTATT-----GTAGTAACTCTAAAGAAAT 6363
Db 1767 CysIleProSerAspThrLysSerGlyAlaGlyAlaThrThrLysSerGlySerAsp 1786
QY 6364 GAAATATGCGATATGATCCCTCTACAGAGAAAAATATGATTAATAATATATATATAT 6423
Db 1787 SerGlySerIleCysIleProProArgArgArgLeuTyrValGlyLysLeuGln--- 1805
QY 6424 TTAATATATGCAACT-----GAAATATAGCGTGAC 6453
Db 1806 ---GluThrAlaThrAlaLeuProGlnGlyGlyAlaAlaProSerHisSerArgAla 1824
QY 6454 AATGATATATAAGAGCTTTTATTAATGTCAGCAATAGAACTCAATTTTGTGTGA 6513
Db 1825 AspAspLeuArgAsnAlaPheIleGlnSerAlaAlaIleGluThrPhePheLeuTyrPhe 1844
QY 6514 AAATATATATATGAA---AATCTGACAGCAAAATGAA----- 6549
Db 1845 ArgTyrLysGlnGlyLysLysProGlnGlyAspGlySerGlnGlnAlaLeuSerGlnLeu 1864
QY 6550 -----TTGCAAAATGCAACA 6564
Db 1865 ThrSerThrTyrSerAspAspGlnGlyAlaAspProProAspLysLeuGlnAsnGlyLys 1884
QY 6565 ATTCCAGATGAAATTTAAAGAAATATGATATATATATGATGATATTAAGATTTGTT 6624
Db 1885 IleProProAspPheLeuAlaGlyLeuMetPheTyrThrLeuGlyAspTyrArgAspIleLeu 1904
QY 6625 TTTGGACATGATTTTCTTAATGATTAATAAAATATATATATATATATATATATAT 6684
Db 1905 ValHisGlyGlyAsnThrSerAspSerGlyAsnThrAsnGlySerAsnAsnAsnIle 1924
QY 6685 ATTCTCAATGAAATATATAGAAATAACAGATTAATAAAATATTAAGATTAACGTA 6744
Db 1925 ValLeuGlnAlaSerGlyLysLysGlnAspMetGlnLysIleGlnLysIleGlnGln 1944
QY 6745 ATA----- 6747
Db 1945 IleLeuProLysAsnGlyGlyThrProLeuValProLysSerSerAlaGlnThrProAsp 1964
QY 6748 ---TTTGGAGCAAAAAATTAATAATTTATTTGGAGAGAAATGATATATGATTAAC 6804
Db 1965 LysTyrPheAsnGlnHisAlaGlnSerIleTyrPheGlyMetIleCysAlaLeuThrTyr 1984
QY 6805 -----CATCTCACAGCAAAACGAAAAAGAAATTTAGA 6840
Db 1985 ThrGlnLysAsnProAspThrSerAlaArgGlyAspGlnLysLysIleGlnLysAspAsp 2004
QY 6841 GATAATTAC----- 6849
Db 2005 GluValTyrGlnLysPhePheGlySerThrAlaAspLysHisGlyThrAlaSerThrPro 2024
QY 6850 -----CAGTACAAAT---GACATGACCAAACTG----- 6873
Db 2025 ThrGlyThrTyrLysThrGlnTyrAspTyrGlnLysValLysLeuGlnAspThrSerGly 2044
QY 6874 -----ACGCTTCCTCCCTTGAAAGCTTTGTAAGAGCC 6906
Db 2045 AlaLysThrProSerAlaSerSerAspThrProLeuLeuSerAspPheValLeuAspPro 2064
QY 6907 CAATTTTGATGATGTTCAAGAAATGGCAGAGAAATTTTGTATTAATGAAGAGAAAGAG 6966
Db 2065 ProTyrPheArgTyrLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2084
QY 6967 TTTGTTAAATTTGAGCGCGGCTGTGAAGAAATATGAG----- 7002
Db 2085 LeuAlaGlnIleLysHisGlnCysLysValGlnGlnLysAsnGlyGlyLysSerArgArgGly 2104
QY 7003 -----TGTAAAT-----GCTAGTAAAT 7017
Db 2105 GlyIleThrArgGlnTyrSerGlyAspGlyGlyAlaLysAsnGlnLysLeuProLysAsn 2124
QY 7018 GACGTAAGACACACAGAA-----TGTGACAGAGCGGTGTATACATATCA 7062
Db 2125 AspGlyThrValProAspLeuGlnLysProSerCysAlaLysProCysSerSerTyrArg 2144
QY 7063 AATTTTATTTAGAGTGAAGAACTGAATATGAAGAAACAGAAAGAAAGTTCAAAAAGAT 7122
Db 2145 LysTyrPheIleGlnSerLysGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2164
QY 7123 AAAGAT----- 7128
Db 2165 LysAspLysCysValAsnGlySerAsnLysHisAspAsnGlyPheCysGlnThrLeuThr 2184
QY 7129 ---GGCAAAAAGATATAGATATAT-----CCTTCTACTGAAAGACATAT 7170
Db 2185 ThrSerSerLysAlaLysAspPheLeuLysThrLeuGlnProCysLysProAsnAsnVal 2204
QY 7171 GACAGGCAACATGCTCCTCATGATATTAACATGAATTAAGAAATATATCTGCAAT 7230
Db 2205 GlnGlyLysThrIlePheAspAspAspLysThrPheLys-----HisThr 2219
QY 7231 AAGGATGTTCTGTGATGCAAAAACCTTCTCTCAACATACCAAAAACACACAATCA 7290
Db 2220 LysAspCysAspProCysLeuLysPheSerValAsnCysLysLys----- 2234
QY 7291 CAATCATCCGATGCTATATGATATGACCAATCGCTGATATATGCTCAAGAAATTAAC 7350
Db 2235 -----Asp 2235
QY 7351 AAGTGTAGTCTCTCAACTTTCAAAAGAGATCTATGATGATCTACAAAAGAAATTA 7410
Db 2236 GluCysAsp-----AsnSerLysGlyThrAspCysArgAsnLys----- 2248
QY 7411 GAACCTAAATATACCTATGATTTGTAAGAAAGACATATATATATATATATATAT 7470
Db 2249 -----AsnSerIleAsp-----AlaThrAspIle 2256
QY 7471 GAAATATATGATATATATCTTGAAGAGAAATTTATATCTATGATGATCTACAAAAGAA 7530
Db 2257 GluAsnGlyValAspSerThrVal-----LeuGlnMetArgValSer 2270
QY 7531 AACGAAGTAATAATAGTGGACATTAATATATCTTGGATCTTAAGAAACCTTAATGA 7590
Db 2271 AlaAspSerLysSerGlyPheAsnGlyAsp----- 2280
QY 7591 CCGATTAATATATATGAGAAAGAAACCTTGAAGAAAT----- 7629
Db 2281 -----GlyLeuGlnAsnAlaCysArgGlyAlaGlyIlePheGlnGlyIle 2295

OY	4102	AAAGAGCAAAATTCCTCCCGATTCTTGGATGATTCACATCCATGCTTACACATTTGGAGATTAT	4101
Db	1445	-----GlyLeuIleProSerGlnPheLeuIvsSerMetCtyTrpPheGlyAspIyr	1462
OY	4162	AGAGATTCTTTTATTTGGACACGATATATATCAAAAGCATAGTCAGAGCAAGTAACCTAAA	4221
Db	1463	ArgAspIleCysLeuAsenThrAspIleSerLysGlnAsnAspValAlaLeuAlaLys	1482
OY	4222	GAGCAATATGATTCTCTTTCAAAAATGGTGACCAAAAATCTCCTAATAGAAAAACACC	4281
Db	1483	AspLysIleGlyLysPhePheSerLysAspGlySerLysSerProSerClyLeuSerArg	1502
OY	4282	CAGAAATGGCGACAGAAACATCTCATGTGATATGGGACATATGCTATMGCACTACTA	4341
Db	1503	GlnGutTrpIlePyrLysThrAsnGlyProGlnIleTrpLysGlyMetCysAlaLeuThr	1522
OY	4342	AAAAATGGCGCAAAAAGATGAT-----TTTACCGAAACATACGCTTACACACAC	4392
Db	1523	LysTrpValThrAspThrAspAsnLysArgLysIleLysAsnAspTyrSerTyrAspLys	1542
OY	4393	GTCAATTATAGTGACAAA---ACACCACTTTGGAGAAATTTGCCAAACCCCACTT	4449
Db	1543	ValAsnGlnSerGlnAsnGlyAsnProSerLeuGlnGluPheAlaLysProGlnPhe	1562
OY	4450	TTAGATGGCGCAACCGAAGTGTACGACGACTATGCTATACACGCAAAAATATTGAC	4509
Db	1563	LeuArgTrpMetIleGlnTrpGlyGlnGluPheCysAlaGlnGluIleLysGlnAsn	1582
OY	4510	GATGTGCAGAAAAATGTAACTCA---AATGACCAATGAAAGTGTATACA-----	4557
Db	1583	IleLeuLysAspAlaCysAsnGlnIleAsnSerThrGlnGlyCysAsnAspAlaLysHis	1602
OY	4558	GATGTATATAGAAATACGAGGACTCGTGTAAATATATGAAA---AAAAAAAAGAGTG	4614
Db	1603	ArgGlyAsnGlnAlaCysAlaGalaTyrGlnGluTrpValGlnAsnLysLysGluPhe	1622
OY	4615	ATTCACACAGATTAATATTTACAAGATGAACCGCAAAAAAAGATTCATAGACACAC	4674
Db	1623	SerGlyGlnThrAsnAsnPheValLeuLysAlaAsnValGlnProGlnAspProGluLyr	1642
OY	4675	ATTTGGTGAATCGTTACAGACTTACTGGACAGCAATGCACAACGATTACTTGACACGAA	4734
Db	1643	LysGlyLyrGluTyrLysAsp-----	1649
OY	4735	TTTACTGCTAGTTGGTGTGATACCCCTGGAGAGTGCCTGTGATCAAGAATATACA	4794
Db	1650	-----GlyValGlnPro-----IleGlnGlyAsnGluLyr	1659
OY	4795	TTGTAGAAAAACAGCGCTTACTATCTATGCCGACAAACAATTGGGGTCACAAAATTTAT	4854
Db	1660	LeuLeuGlnLys-----CysAspAsnAsnLys--CysSerCys-----	1671
OY	4855	GAATATGACGCAAAATATCTAATCAATTTGAGTAAAGATTAAGTGCAAAAGATTAGTAAG	4914
Db	1671	-----	1671
OY	4915	GAGCAAAACACAGGTGCTATTAAATGGCAAAACAAAGTCTAATTAATTAATCTTG	4974
Db	1671	-----	1671
OY	4975	AAAGATTTGACTGAAGATGTGCTTTTCCCTTCGTCGACTACGTATATGTTTCATGCA	5034
Db	1671	-----	1671
OY	5035	TTTGATGGCAATTATACAGATCCAGAAATTAAAGATGAATAATGGGTGGCAAAAAGATTG	5094
Db	1672	MetAspGlyAsnValLeuSerValSerProLysGluLys-----	1684
OY	5095	ATGGAATGGCGCGCAACGGAAGGTATCAATTTGGTGCATCTACTACAAAAGAAAAAGAA	5154
Db	1685	-----ProPheGlyLysTyr-----	1689

QY	5125	AAAGGAAAAAATAAAGCTGGAGTGGCGACAAATATCTTATAGAGTCCCGCTGTAGT	5124
Db	1690	----- -----AlaHisLysTyrPro-----	1694
QY	5215	GCTATGAATATAGTTTTTATGATTTAAGAGATATATCTAGATTGATTAATTTGGAA	5274
Db	1694	-----	1694
QY	5275	GATGAAAAACAAAAGACCGAGAAAATTGAAGAAAATTTTAACAAAAATGGAACATCA	5334
Db	1694	-----	1694
QY	5335	GTTGGCAAAAGAAAGTGAATAGTACTACAGAAATCCCGTAGTACTCCGAAAAATTTTTC	5394
Db	1694	-----	1694
QY	5395	TGGAACGAAATTAAGAAATGTGTGTGGAAACGAATGATATGGGGGTACAAAAGTGTAGG	5454
Db	1694	-----	1694
QY	5455	GATGATGGAATAGTGGAAATAGTGCAGAAGTGAAGATCTAAAAAATGTGGTTCT	5514
Db	1694	-----	1694
QY	5515	GTACCTTCAGATGATGATATATCTATGGGAAAAATCGCGATGAAGTACTGCGTATCAG	5574
Db	1694	-----	1694
QY	5575	TTTCTGATGCTGTTGCCGAATGGGGGTAGAGATTTTTCGAAACATTAAGAAAGAAATG	5634
Db	1694	-----	1694
QY	5635	GAGAAATTTGTAGGGGCGGTGATGATTAATACTTGTGTATATGAAGATTAAGAAG	5694
Db	1695	----- ::: -----Glu	1695
QY	5695	AAATGTACAGATCGCTGTACAAATATAAAAAATTTATAGTGAAGAACACAGAT	5754
Db	1696	 LysCys-----	1697
QY	5755	GAAAAACAAATCABAAAAATATGTGAGAAATTAAGACAAAATATATTCGAGCATCTGTG	5814
Db	1697	-----	1697
QY	5815	GCAAAAGATCGAGAGAGACGCTCGGAATATTAGACAAACAATTAATAAATTTGTGA	5874
Db	1697	-----	1697
QY	5875	AATAAACTGAGATTGTGAATATAAGTGTATGAAAGATGTGTCCACACGCGATTAACT	5934
Db	1698	 -AspCys-----	1699
QY	5935	GATGTTATAGTCAAAATATGCGCGCATCTTAGACAGATGAACCAAAAGAGTTAGGA	5994
Db	1700	----- ::: -----TyrGlnGly	1702
QY	5995	AAAGTAAATTTCAAGTGGCCACGAGGTCCACACGTTGACGAAGGAAACACCGTCACCA	6054
Db	1703	 Lys-----HisValProSerIleProPro-----	1710
QY	6055	CGGGTATCATGTATATCAAAAGGACGCGATCGAAAAAAGACGAAAACAGCGCGCT	6114
Db	1711	----- -----ProProPro	1713
QY	6115	ACAAAACGCCGAAAAAATGTGAAATCTTAACAACGAAATGCGAGCACAAACAGAAC	6174
Db	1714	 ProValGlnProGln-----	1718
QY	6175	CGAGAGACAGACAACAACACGAAAACGAATCAACAGACAACAACAGAAATCTGAC	6234
Db	1719	----- ProGlnIleProThrValThrValAsp	1727
QY	6235	GTGGGCAATATGTTAAAGCCATCTTTCGATTAACCAAGATAGCAGGGGTGGAATAGG	6294

```

Db 1728 ValCysSerIleValIysThrLeuPheLys-----AspThrAsnAsnPheSerAsp 1744
QY 6295 GGTTGTAATCCAAAACAGATGACAAATAT-----CCTAAAGGGGTGTATAT----- 6342
Db 1745 AlacysGlyLeuLys-----TyrGlyLysThrAlaProSerSerTyrLysCysIleProSer 1763
QY 6343 -----GTAGCTAAGCTCTAAGAAATGAAATGAAATGCAATG 6375
Db 1764 AspThrLysSerGlyAlaGlyAlaThrThrGlyLysSerGlySerAspSerGlySerIle 1783
QY 6376 TGTATGCCCTGCTAGGAGAAATATATGTAATATATATATATATATATATATATATATAT 6435
Db 1784 CysIleProProAlaGlyArgGlyValGlyLysLeuGln-----GluTyrAla 1801
QY 6436 ACT-----GAAATATAGCTGACATGATATATATA 6465
Db 1802 ThrAlaLeuProGlnGlyGlyAlaAlaProSerHisSerTyrGlyAlaSerPheLeuArg 1821
QY 6466 GAGGCTTTTATTAATGTCAGCAATAGAACTCAATTTTGTGTTAAATATATATATAT 6525
Db 1822 AsnAlaPheIleGlnSerAlaAlaIleGlnThrPhePheLeuThrPsrPrgTyrLysGlu 1841
QY 6526 GAA---AATCGCTCAGCAGAAATGAA----- 6549
Db 1842 GluLysLysProGlnGlnLysAspGlySerGlnAlaLeuSerGlnLeuThrSerThrTyr 1861
QY 6550 -----TTGCAAAATGACAATTCAGATGAA 6576
Db 1862 SerAspAspGlnGlnLysProProAspLysLeuLeuGlnAsnGlyLysIleProProAsp 1881
QY 6577 TTTAAAGAAATATGTAATATACATATGCGTATTAAGATATGTTTGTGAACTGAT 6636
Db 1882 PheLeuArgLeuPheThrLeuGlnLysAspTyrArgAspIleLeuValHisGlyGly 1901
QY 6637 ATTCTAATGATAAAAAATTAATTAACCTGTAACAATAGTGAACAACCTGTCATGAA 6696
Db 1902 AsnThrSerAspSerGlyAsnThrAsnGlySerAsnAsnAsnIleValLeuGluAla 1921
QY 6697 AATATTAAGAAAAACAGATTAATAAATAAGATTAAGAAATTAACGTAATA 6747
Db 1922 SerGlyAsnLysGlnAspMetGlnLysIleGlnGlnLysIleGlnLysProLys 1941
QY 6748 -----TTTGGGAG 6756
Db 1942 AsnGlyGlyThrProLeuValProLysSerSerAlaGlnThrProAspLysTyrTrpAsn 1961
QY 6757 AAAAATAAAAAATTAATTTGGAGAGAAATGATATATGATTAACCTAT 6804
Db 1962 GlnHisAlaGlnSerIleTyrPlyGlyMetIleCysAlaLeuThrTyrThrGlnLysAsn 1981
QY 6805 -----CATCTCAGACGAAAAACGAAAAAATTAAGATTAATTAAC 6849
Db 1982 ProAspThrSerAlaArgGlyAspGlnLysIleGlnLysAspAspGlnValTyrGln 2001
QY 6849 ----- 6849
Db 2002 LysPhePheGlySerThrAlaAspLysHisGlyThrAlaSerThrProThrGlyThrTyr 2021
QY 6850 -----CAGTACAAT--GACATGACCAACAG 6873
Db 2022 LysThrGlnTyrAspTyrGlnLysValLysLeuGlnLysAspThrSerGlyAlaLysThrPro 2041
QY 6874 -----ACCCCTTCCCTTGAAGAGTTTGTAAAGAGCCCAATTTTGTGA 6918
Db 2042 SerAlaSerSerAspThrProLeuLeuSerAspPheValLeuArgProProTyrPheArg 2061
QY 6919 TGGTTACAGAGATGGGACAGAAATTTGTAATGAAGAGAGAGAGAGAGAGAGAGAGAG 6978
Db 2062 TyrLeuGlnGlnLysTyrGlnLysAsnPheCysLysLysArgGlyHisLysLeuAlaGlnIle 2081
QY 6979 GAGCGGCGCTGAAGGAATATAGC----- 7002

```

```

Db 2082 LysHisGlnCysLysValGlnGlnLysAsnGlnGlyGlySerArgArgGlyGlyIleThrArg 2101
QY 7003 -----TGTAAT-----GGTAGTAATGAGGTAAAGACA 7029
Db 2102 GlnTyrSerGlyAspGlyGlnAlaCysAsnGlnMetLeuProLysAsnAspGlyThrVal 2121
QY 7030 CAAGAA-----TGTCCAGAGCGCGTGTAACTATCAAAATTTTATTAAG 7074
Db 2122 ProAspLeuGlnLysProSerCysAlaLysProAsnAsnValGlnGlyLysThrIleGln 2141
QY 7075 AAGTGAAAACTGAATATGCAAGACAAAGAGAAAGTTCAAAAAGAGATAAGAT 7128
Db 2142 SerLysGlyLysGlnPheGlnLysGlnLysLysAlaTyrGlnGlnGlnLysAspLysCys 2161
QY 7129 -----GCGAAAAAG 7137
Db 2162 ValAsnGlySerAsnLysHisAspAsnGlyPheCysGlnThrLeuThrThrSerSerLys 2181
QY 7138 TATAAGCAATAT-----CCTTCTACTGAAAGACATGACAGAGACACA 7182
Db 2182 AlalysAspPheLeuLysThrLeuGlyProCysLysProAsnAsnValGlnGlyLysThr 2201
QY 7183 TGTGCTCATGAATATTAACATGAATTAAGAAATTAATGTCGAATTAAGATGTCT 7242
Db 2202 IlePheAspAspAspLysThrPheLys-----HisThrLysAspCysAsp 2216
QY 7243 TGTATGCAAAAACCTTCTTCACAACTACCAAAAACACACAAATCAATCAATCCGAT 7302
Db 2217 ProCysLeuLysPheSerValAsnCysLysLys----- 2227
QY 7303 GCTAATGATATGCCAGAAATCCCTGGATTAATGTCCTGAAGAATTAACAAGTGAAGT 7362
Db 2228 -----AspGlnCysAsp 2231
QY 7363 CCTGAACTTAAAAAGGAGTCTATGATTCATACAAAAAATTAAGTGAACCTAAATA 7422
Db 2232 -----AsnSerLysGlyThrAspCysArgAsnLys----- 2241
QY 7423 CCTATGAATTTGTAGAGAAAGACAGCATATTTATTAAGAGAGCAAAATTAATG 7482
Db 2242 -----AsnSerIleAsp-----AlaThrAspIleGlnAsnGlyVal 2253
QY 7483 GATATTAACCTTGAAGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7542
Db 2254 AspSerThrVal-----LeuGlnLysValSerAlaAspSerLys 2267
QY 7543 AATAGTGGAACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7602
Db 2268 SerLysPheAsnGlyAsp----- 2273
QY 7603 ATAGAGAAAGAAACCTTGTGAATAT-----AGAGAAAGAAAT 7641
Db 2274 -----GlyLeuGlnLysAlaCysArgGlyAlaGlyIlePheGlnLysIleArgLysAsp 2291
QY 7642 CGTTTAAAGTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7701
Db 2292 -----GluTyrLysCys--ArgAsnValCysGlyTyrVal 2302
QY 7702 AAAAGATATGTGAACTCCAGAGAGAAACATATGTGCTTAAGAAATTTAGATGAAT 7761
Db 2303 -----ValLysLysPro----- 2306
QY 7762 AAAATGAAGAACTTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7821
Db 2306 ----- 2306
QY 7822 AATGAAGAAATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7881
Db 2307 -----GluAsnValAsnGlyGln----- 2312
QY 7882 ATATGTGATCACTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7941
Db 2313 -----AlaLysGlyLysHis 2317

```

```

QY 7942 ATGTACGAATGGTGGTACTACCTCCGAGAGAAATAAATTAAGTTTGA 8001
Db 2318 IletlelInleatgalaleuVallys-----ArgtlrValgluYrPhePhe 2333
QY 8002 TACATATATGAAATGAGAAATATAAAGTAGAATAATATCAACGAT----- 8055
Db 2334 GluAspTyrAsnLysIleLysHisLysIleSerHisLysIleSerHisLysIleSer 2353
QY 8056 -----CTACAAACGTTTCTGCTTGGTGGATGCTATATGAAAGATATTGGAAA 8109
Db 2354 ProcysIleLysAsnGlyValGluLysTTPValAspGlnLysAsnGlyGlu---TTPlys 2372
QY 8110 GCAATGACGTGCAAGCAGCAGAGATGCAAACTTTTGAAGAGAGATGATGA 8169
Db 2373 GluIlethrgluArpPheLysAsp----- 2380
QY 8170 TTTGAAACGCAATTAATATCAAGATTAAGTGGACATTAAGACATCCACTTTGAT 8229
Db 2381 -----GlnTyrLysAsnAspAsnSerAspAsnValArg 2393
QY 8230 GATTATATACCTCAGCGTTTTCATGATGATGATGATGATGATGATGATGATGATGAT 8289
Db 2394 SerPheLeuLurhLeuIlePrgInIlethAspAlasn-----AlalysAsn 2410
QY 8290 CTGATGAGAGAAATGAGAAATTTAAATCATGTGATGATGATGATGATGATGATGATGAT 8349
Db 2411 LysValIleLysLeuSerLysPheGlyAsnSerCysGly---CysSerAlaSer----- 2427
QY 8350 TGCAGAAATATATATGATGAAATTAAGTGTGACAGTGTGAAACGATGTCAAGATAT 8409
Db 2428 ---AlasInLugInAsnLysAsn-----GlyGluTyr 2437
QY 8410 AAAATTTTCTTCTTAATGAAATCTCTATGATATCAATCAATTAATATCAAGAA 8469
Db 2438 LysAspAlaIle-----AspCysMetLeuLysLysLeuLysAsp 2450
QY 8470 TTGTATGACACACCAATATATACAAAATCTCTATGATGATGATGATGATGATGATGAT 8529
Db 2451 -----LysIleGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2460
QY 8530 CAAAAGTTGAAACTTTTAAAGGATGCTGCTGAGAGCTTTTGTGATATCTTCAT 8589
Db 2461 -----GlnThrSerAspThrGluCysSerAspThrProGlnProGlnThrLeuGlu 2477
QY 8590 GAAACAGTAAGTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8649
Db 2478 Asp-----GluThrLeuAspAspAspIleGluThrGluLysLysLysLysLysLysLys 2495
QY 8650 ACATATGCTTTTCGAGAGAACACCAAAAGTTATTAAGAA-----GCTTGAGAT 8697
Db 2496 ProLysIleCysGluAsnValLysLysThrAlaGlnGlnLysLysLysLysLysLysLys 2514
QY 8698 TGTACACTCTCTTAAGATTCATTTGATGATATGCTTCCATCAAAACAAAGAT 8754
Db 2515 -----ValProAlaGluAsnSerGluProAlaIleThrAspSerGlyLysGlu 2531

```

```

RESULT 8
AAB18352
ID AAB18352 standard; protein; 2197 AA.
AC AAB18352;
XX 07-NOV-2000 (first entry)
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:210.
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
OS antimalarial; malaria; protozoa; infection; insecticide.
XX Plasmodium falciparum.
XX NO200025728-A2.

```

```

XX 11-MAY-2000.
PD 11-MAY-2000.
XX 05-NOV-1999; 99WO-US26796.
PF 05-NOV-1999; 99WO-US26796.
XX 05-NOV-1998; 98US-0107131.
PR 05-NOV-1998; 98US-0107131.
PA (HOPE/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
PI Hoffman S, Carucci D, Gardner M, Venter JC;
DR WPI; 2000-365347/31.
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
XX diagnosis of P.falciparum infection.
XX Disclosure; Page 441-447; 577pp; English.
XX The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II).
XX (I) and (II) are useful for the development of vaccines against
XX P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX antibody raised to immunogens comprising the sequences of (I), are
XX useful in the detection of infection with P. falciparum. Furthermore,
XX (I) (especially when they are rifins or secreted or membrane proteins)
XX can aid the identification of drugs to treat or prevent P. falciparum
XX infection, or they can be used to identify drug resistance in
XX P. falciparum. Sequencing of the plasmodium chromosome 2 and the
XX subsequent identification of proteins encoded by it will help to expand
XX our understanding of parasite biology, a process hampered by the
XX complexity of the parasite life cycle, and provide new targets for
XX vaccine and drug development. Parasite resistance to drugs and mosquito
XX resistance to insecticides have led to a resurgence of malaria in many
XX parts of the world, and there is a pressing need for vaccines and new
XX drugs. AAT70078 to AAT70287 and AAB18144 to AAB18352 represent nucleotide
XX and protein sequences given in the present invention, but which are not
XX specifically mentioned within the specification.
XX
XX Sequence 2197 AA:
SQ
XX
XX Alignment Scores:
XX Pred. No.: 2,99e-116 Length: 2197
XX Score: 1773.00 Matches: 761
XX Percent Similarity: 29.27% Conserved: 323
XX Best Local Similarity: 20.55% Mismatches: 735
XX Query Match: 9.11% Indels: 1885
XX DB: 21 Gaps: 120
XX
XX US-10-087-013-1 (1-10628) x AAB18352 (1-2197)
QY 97 AAAAGTGCAGAAATGTTTGAACGTTATGCCAAAATATTAAGATCCATCAAAA--- 153
Db 12 GluSerValLysHisMetPheAspArgIleGluAspValLysGluInValLysSer 31
QY 154 TATGCAAAAGACATGTGATGCTTGAAGAGGAGATTGACGAAAGACGAAATTCGTGT 213
Db 32 GluThrValAsnTyrValSerGluLeuGlnGlyLysLeuSerLeuAlaProIleLeuGly 51
QY 214 GGTCCCTTACGCGCAGTAATTAAGCAATTAATTAATTAATTAATTAATTAATTAATTAAT 273
Db 52 ValGluSerLysSerThrAsnGlu-----ThrCysAsnLeuVal 64
QY 274 CATTAAGACATATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 333
Db 65 GluAspTyrTyrAsnLysProValLysGlyAsnSerAsn---ArgTyrProCysLysAsn 83
QY 334 -----AGAGAAACAAACGATTTGATGATGATGATGATGATGATGATGATGATGAT 381

```

```
Db 84 LeuLysGlyIleThrAsnGluArgPheSerAspThrLeuGlyGlnCysThrAsn 103
QY 382 AAAATA-----CGTAATTATATAAAAGAAAAAT-----CATGCTATAGCTGCGC 426
Db 104 LysLysIleLysGlnGluThrSerThrLysSerLysAspCysGlyAlaCysAla 123
QY 427 CCACCTGAAAGACGACATATGTGTGATATAAAACTTGAGACGCTGAATGATATAATACC 486
Db 124 ProTyrArgArgLeuIleLysCysSerHisAsnLeuGluSerIleAsp-----ThrThr 141
QY 487 CAAAATTCATGATTATATGGGAAATAGTACTAGTACACCAAAATTCGAGGTGAATCA 546
Db 142 SerMetThrHisLysLeuLeuGluValCysMetAlaIleLysTyrGluGlnLysSer 161
QY 547 ATTGTAATATCATCA-----CATAAAGAACT-----TCAGACGGCTGT 588
Db 162 IleAspThrHisTyrProGlnHisGlnArgThrAsnGluAspSerProSerGlnIleCys 181
QY 589 ACTGCTCTTCACGAACTTTCGACATATAGTGATATGTAAGAAATAGATATGTT 648
Db 182 ThrMetLeuAlaArgSerPheAlaAspIleGlyAspIleValArgGlyLysAspLeuPhe 201
QY 649 AAACCAAT-----GTCCATGACAAAGTAAACGGGCTCCGAGAGCTT 693
Db 202 TyrGlyAsnSerLysGlnLysGlnLysArgAspGlnLeuGlnThrAsnLeuLysThrIle 221
QY 694 TTCAGAAATATACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 747
Db 222 PheGlyLysIleHisGlnLysLeuLysAspLysGlnGlyAlaGlnThrArgTyrGlySer 241
QY 748 GATGATCTCGGAAATATATATAATTAAGAAACATGCTGTAATGTAAGAAATAA 807
Db 242 Asp---ThrThrasnTyrGlnLeuArgGlnAspTyrPyrAlaAsnArgAlaThr 260
QY 808 GATGGAAGCTATACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 864
Db 261 ValTyrPheIleAlaIleThrCysAspVal---HisGlySerAspTyrPheArgGlnThrCys 279
QY 865 -----GAAAGTAATACACCATTAATTTTCAATTCCTAAATGC----- 900
Db 280 GlyAspLysGlnIleThrAlaThrArgValLysAspLysCysAlaArgCysLysAspGluAsn 299
QY 901 GGGCATAAACAAGCA-----AAGTCTCCACCAATTTAGATATATGCTCCCTCA 948
Db 300 GlyLysLysProGlySerAsnAlaAspGlnValProThrTyrPheAspTyrAlaProGln 319
QY 949 TATTTACGTTGGTTCGACGATGCGGAGAGAGATTGGCGGAAATATTAATTTG 1008
Db 320 TyrLeuArgTyrPheGlnGluIleThrAlaGlnAspPheCysArgLysLysLysLeu 339
QY 1009 AAAAAGTCAAGAGACTCCTGCTGTAATGACAAAGAACGCTTAATTTAGTCAATGCA 1068
Db 340 GlnLysLeuGlnGlnGlnCysArgAspTyrLysGlnAsnLeuTyrCysSerGlyAsnGly 359
QY 1069 CATGATTCGACGACAACTTTGGGAAAAAGTATTTGCACTTTGATGATTAAGTACT 1128
Db 360 TyrAspCysThrLysThrIleTyrLysLysGlyLysLeuValIleGlyClnHisCysThr 379
QY 1129 GACTGTTGACATAATGCAAAAGTTTGGATTTGTTAGGAAATCAACAAAGACATTT 1188
Db 380 AsnGlySerValTyrCysArgLeuTyrGlnSerTyrPheAspHisGlnLysLeuGluPhe 399
QY 1189 AAAAAACAAGAAAAATATGAAAAAGAAATACATATATTTATTCG----- 1236
Db 400 LeuLysGlnLysGlnLysTyrGlnIleThrGlnIleSerAsnSerGlySerCysGlyLysSer 419
QY 1237 -----ACGATACAAATTTGTCATATATATATTAATAGTAA----- 1272
Db 420 GlyGlyValLysGlyArgAsnArgLysArgGlyAlaGlyValGlnThrAlaThrAsn 439
QY 1273 -----TATTAATAACATTTTATGAAAAAATTAAAGAAACGACATATGCAACTAAT 1323
Db 440 TyrAspGlyTyrGlnLysLysPheTyrLysGlnLeuLysGlnSerGlnTyrGlyLysVal 459
QY 1324 GACACTTTTAAATTTACTAAATGAAGAAAGTATGTAAGAGAGATTAACAGAGAA 1383
Db 460 AspAspPheLeuLysLeuLeuAsnAsnGlnAspValCysLysLysIleLysAspGlnLys 479
QY 1384 AAGGATATTAATCTTTAAAGAGTGTGAT-----GACAAAGGATATTTAT 1431
Db 480 GlnLysIleAspPheThrLysProAlaAspLysAsnSerAsnAsnGlnIleThrPheTyr 499
QY 1432 CGTTGCAAAATTTGCGCAACTGTGTCGCGACCTCGGGGTCAAAATGATGATGTAATAATAC 1491
Db 500 HisSerGlnTyrCysLysProCysProAspCysGlyValLysArgLysAspAsnGlnTyr 519
QY 1492 ACACCAATTCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1551
Db 520 LysAspLysTyrAspLysCysThrArg-----GlyLysLeuTyrGlnProAlaSer 537
QY 1552 GGTGTGAACCTTACTAAATTTACTGCTCTTATGTTAGTAAAGCAAGGATATTTACA 1611
Db 538 GlyAlaGlnGlyThrProIleLysIleLeuLysSerGlyGlnLysGlnLysGlnIleGlu 557
QY 1612 CAAAAATTCAAAAATTTTGTAAAC-----ACCTCAACTAAT----- 1647
Db 558 ThrLysLeuLysAlaPheCysAspGlnThrAsnGlnLysAspThrThrAsnSerValAlaArg 577
QY 1648 -----TACAAACATATAAATATATCAAAA-----TGG 1674
Db 578 GlyGlyGlyAlaAspGlySerGlySerLysSerAsnSerLysGlnLeuTyrGlnGluTyr 597
QY 1675 GAATGCTATTTAAAG----- 1689
Db 598 LysCysTyrAsnGlnValGlnLysValLysAspLysAsnGlyGlnGlnLysAspGlu 617
QY 1690 ---GATGAAATATAAATAGA-----TGTAACGTGAA----- 1719
Db 618 AspGlnGluAspValAspLysValLysAlaGlyLysLeuLysIleLeuGlnAsnLys 637
QY 1720 -----CAAAATACGAAATCAATTAATGATAATCTTAAGATATATCTATTGAT 1767
Db 638 LysHisGlnSerArgAsnAsnSerSerAsnGlnProGlnIlePheGlnLysThrPheHis 657
QY 1768 AATTTTGTGATATAGGTTATATTTAAAGGATATATTAAGATGACAA 1827
Db 658 AspPhePheTyrPheTyrPheIleGlyArgPheLeuAsnAspSerMetTyrTyrPargGlyLys 677
QY 1828 CTAAAACTGTATATAAATAT---ACAACACGCAATGATGATGATGATGATGATGATGAT 1884
Db 678 ValAsnSerCysIleAsnAsnProLysArgLysLysCysArgAsnGlnCysLysAspAsp 697
QY 1885 TGCATTATGTTTGAACAGATGGGTTAAACAAAGAAAGAAAGATGATGATGATGATGAT 1944
Db 698 CysGlyCysPheLysGlnTyrPheIleGlyLysLysLysGlnGlnIleThrGlnLysLys 717
QY 1945 CTGTTCACAAAAAAAGAAATATACAGCAATCGTATTAATATTAATCTTTT 2004
Db 718 HisPhe----- 719
QY 2005 GAAGGTTATTTTAAAGTTATGATTAACCTTGACAAAGTGAAGCAAAATGCAAGAA 2064
Db 720 -----LysTyrGlnGln 723
QY 2065 CTTATGAAATATATAAAGAAAAAATGAGTTTCCATTTGCAAAATTAATAGGAC 2124
Db 724 AlaPheLysAsn---LysArgGlnAsnSerGlyIleAspMetPheSerGlyLeuMetAsp 742
QY 2125 TATTTAAGAAATGCAATAGAACTCTTGTAGATCACTTAAAGAAACTGCCAGATATGT 2184
Db 743 SerAlaAspValValIleGlnLeuValAlaLeu----- 752
QY 2185 AAGACAAATATATCAAAAGCAAGCATGTGAAGCATCCATTAATGACAAACAAACCCGTGT 2244
Db 752 ----- 752
```


[illegible]

OY	7561	AATCCCTTGGCATGCTTAAAGAAACCTTATGCACTGTAATATATAGGAAG	-----AGA	7614	
Db	1361	GLYValLysGLYThrLeuLysAspSerAlaAlaGlu	-----PheLeuAsnHqLeuLysAsn	1379	
OY	7615	AACCCCTGTGAATATAGAGAGAAATCGTTTTTAAGTACATATATGATGAATGGTAC		7674	
Db	1380	GLYProCysLysAsnGluSerGluGluAsnLysLysAlaGluAspGluIleAspPheLys		1399	
OY	7675	AAAAATTCAAAAGTTCTATCAGAGAGAAAAAAGAGTATGTGTACCTCCAGAGAGAACAT		7734	
Db	1400	LysProAspSerPheThrPheLysAspAlaAspAsn	-----CysLysPro	1413	
OY	7735	ATGCGCTTAAAGCAATTTACATGAATTAATTAATGCAAGACTTAAAGATAGTATATCTC		7794	
Db	1414	-----CysSerGluPheLysIleLys		1420	
OY	7795	CTAATAATGGTTCGTGCACTGCAGAACTGCAGAAATGAAAGATATACATATATAAAACCTTCAC		7854	
Db	1421	-----	-----CysGluAsnHisAsn	1425	
OY	7855	TCAGAGAACGGGTGCCCAATGATCCAAATATGTGAT	-----ACTATGAATATAGTTTCGCT	7911	
Db	1426	CysSerSerGlyLysAsnThrGlnGlyLysCysAspGlyLysThrThrIleAlaIleAlaThr		1445	
OY	7912	GATCTGGGTACACTACTAGAGACA	-----GATATGTTACGAATTTGGTGGTAC	7962	
Db	1446	GluIleGluAsnIleLysThrAsnThrIleGluValIleThrMetLeuValSerAspSer		1465	
OY	7963	TTACCTCCGTACAAATTAATAATATATTAAGTTTGAATACATATATGMAATGAGAG		8022	
Db	1466	LysSerAlaThrGluPheLys		1472	
OY	8023	AATTAATAATTAAGATAGAAATATAATATACACGATGTACAAAGTTTCGTCTGTGGTG		8082	
Db	1473	-----AspGlyLeuSerGluCysLysAspLysGlyIlePheLysGlyIle		1487	
OY	8083	GATCTAATGAAAAGATATTGGAAAAGCAATGACGTCCAAAGCACAGAGATGCAAA		8142	
Db	1488	-----ArgLysAspGluThrIleGlyCys		1494	
OY	8143	CTTTTGAAGAAAGAAAGATGATGATGATTTGAACCCATTAATTA	-----ATACAAAGATAC	8199	
Db	1495	-----GlyLysValCysGlyValAspIleCysAsnLeuLysLysLysAspAsn		1510	
OY	8200	TGTGACATAGACAGCATCCACCTGTGATGATATATACCTCAACGTTTCGATGATG		8259	
Db	1511	IleLeuLysLysLeuSerAsp	-----LysLysThrIleIleMetLys	-----GluLeuLeu	1526
OY	8260	ACTGAATGCTGTAATATATATGTTAAACGCATGATGAGAGATTTGAAAAATTTAAAAA		8319	
Db	1527	LysArgThrPheuIleuThrPhe	-----LeuGluAspTryAlaLysLysIleLysHis	1542	
OY	8320	TCAATGATACAGCTAAAAACATCTGACAGATGCAAGATGATATGATGAATAATAGTGT		8379	
Db	1543	LysIleSerHisCysThrLysAsnGlyLys	-----GlySerLysCys	1556	
OY	8380	GAAACAGTGTAAAAAGAGATGTCAGAAATATAAAATTTGTCTTAAATGAATCTCTA		8439	
Db	1557	-----	-----IleLysGlyCysValAspLysTrp	1564	
OY	8440	TTGATATACATCAATAATAATAACAAAGATTTATGAAACCAACAAATATATACAAAAATC		8499	
Db	1564	-----		1564	
OY	8500	TCTACTATGATCATGTCTCAAAATTTTGTACAAAAGTTGAATCTTTTAAAGTGAATGT		8559	
Db	1565	-----	-----ValGlnGluLysLysGluGluThrLysGlnIleLys	1576	
OY	8560	TCTGTGAGAGCTTTTCTGAAATATCTTCATGAACAAAGTAAGTGTTTGAATATATAATT		8619	
Db	1577	-----GluArgPheAsnGlu	-----GlnTryLysSer	1585	
OY	8620	AATTAATAATGATGTTCTTCCAAATATACGACATATATGCTTTGGAAGAAACACAAAAAGT		8679	

Db 1586 LysThrSerAspGluTyrPheAsnValIysSerPheLeuGluThrTrpIleProLys--- 1604
QY 8680 TATAAAGAGGTTGCACATTGTACACTACCTTGAAGAATCGATTGTAATTGTCTTACC 8739
Db 1604 ----- 1604
QY 8740 GATCAAAAACAAAGATGATGATGAATTAACAACCTTTACCTTCCTGCAAGAATGAT 8799
Db 1605 ----- 1605
QY 8800 TATGATATAATCTTGTGTAATTTGGAACGCAATCTTGTCTTAAATGTTGATGATATAC 8859
Db 1611 GlnAspAsn-----ValIleLysLeuSerLysPheGlyAsnSer 1623
QY 8860 AAGGTGATATGATCTCCAGAGAGAGACATTTATGTACAGACCTATACATGCAATG 8919
Db 1624 CysGly-----CysSerAlaSerAlaIleSerTrp 1633
QY 8920 AATTATAGAAAAGGTGATAAGAAATTTTAAAAAAACCTTCTTCTGCTTCAGT 8979
Db 1634 Asn-----GlyAsnGluGlu----- 1638
QY 8980 CAAGCAAAATTTAGTCAAAAATATAATCGGAAGAGAGTTGCTTTGAGCAATG 9039
Db 1638 ----- 1638
QY 9040 AATATAGTTATGACAGATTTTCGATATATAAAGAACTGATATGATGACACTTCA 9099
Db 1639 ----- 1639
QY 9100 TTTATCGAAAAATTAATAAAATATTTTGAACATCAATGAAGCAACGAAAAATCTTAA 9159
Db 1645 IleLysLysLeuGluLysLysLysIleAspGluCysLysArgLysProGlyLysAsn----- 1662
QY 9160 ACATGTTGGAAAAATATATAGACGTACATATGACAGCTATGTTATGATGATATAAAT 9219
Db 1662 ----- 1662
QY 9220 GCTACTTCAAAAAGTAAACATTAGATGAGAGAGTGCTCAATTACCAAGAGTGAAGAACT 9279
Db 1662 ----- 1662
QY 9280 AATGATTTCTTCTGTTGTTAATGATGGAAGCAAGCATGTAGAGAAAGAAACAT 9339
Db 1663 ----- 1663
QY 9340 GTAAGTATTCATTAATAAACAATAATGCTCTGTCACAAAGAGATAATTTGAGCGTCA 9399
Db 1668 ----- 1668
QY 9400 GAATTTATTAAGCAACCTGATGTCAGATGATATTAAGAAATATATTAAGTTGAATATA 9459
Db 1669 GluThrLeuThrHisPro----- 1674
QY 9460 TTGATAAAAAATACATGGAATAATCTAATATATAAATAATTAAGCATTAAGATCAATCT 9519
Db 1674 ----- 1674
QY 9520 TCAGGTAATATAGACAATTAACCATCTGAGAGAAATGTTCACTCATATATAAATCAAAA 9579
Db 1674 ----- 1674
QY 9580 GATTCTCATGCGCTTTGGAGTTAATATATTAATGAATAGTTACGAGAACAAAAAT 9639
Db 1675 ----- 1675
QY 9640 AATGAATAATATGATCAAGAAGTACTAAAAAATATATCTGCTGTTTATAT---TTT 9696
Db 1690 AsnPro-----ValGlyLysGlnHisProSerPheCysProPro 1702
QY 9697 GTTGAAGTGAACAACAAAAATCANGTACTAGATGGAATATATAAAGAGAGAG--- 9753
Db 1702 ----- 1702

Db 1703 ValGluAspLysLys-----LysGluGluGluGly 1712
QY 9754 CAACAGTTGTCCTTAACACTGATTTCTTACACCCCAAGTAGATCTTTATCA 9813
Db 1713 GluThrCysThrProAlaSer-----ProAlaProAlaProAlaPro 1726
QY 9814 GCACCTTTTTCACACATGACAGTACGACATATGATCTTAAAAATGAT-----ATA 9867
Db 1727 AlaProAlaSerProSerProThrProAlaProAlaAspGluProPheAspProThrIle 1746
QY 9868 TTGAAGAAGTATCTCTGTTGTTATTTGATCGGCTTGGTGGTATGAGCGTTGATTTTC 9927
Db 1747 LeuGlnThrThrIleProLeuGlyIleAlaLeuAlaLeuGlySerIleAlaPheLeuPhe 1766
QY 9928 ATGAG----- 9933
Db 1767 LeuLysValIleTyrIleCysValValTyrMetTyrIleTyrMetCysPheCysIleTyr 1786
QY 9934 -----AAAAATTCAAATCGTCGTGACCTTGTGCTATACGATGATAATCCGCAAGA 9987
Db 1787 MetTyrLysLysThrLysHisProValAspLeuPheSerValIleAsnIleProLysSer 1806
QY 9988 GAGTATGAGATGCTACGTTGGATCCAAAAATAGTACATACCATATAGAAGTGTCCA 10047
Db 1807 AspTyrAspIleProThrLysLeuSerProAsnArgTyrIleProTyrThrSerGlyLys 1826
QY 10048 TTTAAAGCAAAACATATATATATATGAGAGAGATCT-----AGTGAAGATCA 10098
Db 1827 TyrArgGlyLysArgTyrIleTyrLeuGluLysAspSerGlyThrAspSerGlyTyrThr 1846
QY 10099 GATTAATATATGTTGGAGCTTATCTCTCGATATTTACT-----TCATCGGAAGAGAGAT 10155
Db 1847 AsnHisTyr-----SerAspIleThrSerSerGlnSerGluTyr 1860
QY 10156 GAAGATGATGATATATATATATATATGATGATGATGATGATGATGATGATGATGAT 10215
Db 1861 GluGluMetAspIleAsnAspIleTyrValProGlySerProLysTyrLysThrLeuIle 1880
QY 10216 GAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10269
Db 1881 GluValValIleGluProSerGlyAsnAsnThrThrAlaSerAspThrGlnAsnAspIle 1900
QY 10270 ---AATGATACACACAGTACGAATGATGATGATGATGATGATGATGATGATGATGAT 10326
Db 1901 GlnAsnAspGlyIleProSerAsnLysPheSerAspAsnGluTrpAsnThrLeuLysAsp 1920
QY 10327 GATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10380
Db 1921 AsnPheIleSerAsnMetLeuGlnAsnGluProLysAspValProAsnAspTyrLysSer 1940
QY 10381 GCAGTATTCACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10440
Db 1941 GlyAspIleProPheAsnThrGlnProAsnThrLeuTyrPheAspLysProGluGluLys 1960
QY 10441 CCTTTATATATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10500
Db 1961 ProPheIleThrSerIleHisAspArgAsnLeuAsnGluGluGluTyrSerTyrAsn 1980
QY 10501 ATTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10560
Db 1981 ValAsnMetSerThrAsnSerMetLysAsp---ProLysTyrValSerAsnValTyr 1999
QY 10561 AGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10619
Db 2000 SerGlyIleAspLeuIleAsnAspSerLeuSerGlyAsnLysHisIleAspIleTyrAsp 2019
QY 10620 GAACATATG 10628
Db 2020 GluValLeu 2022
RESULT 9
AAW22476
ID AAW22476 standard; Protein; 2182 AA.


```
QY 3163 ACCGAATGGCGAGATAGTACTGCAAGTGCAGAAAAAGAGTGTAGTAAAGTGAAGGAG 3222
    |||||
Db 844 ThrGlu-----GlnGlu 847
QY 3223 AATGTAAAGAGTGTAAAGATTAAGATATGTCAAGGCTGTACGAAGAGAGTGTACA 3282
    |||||
Db 848 GlyValIleProCys----- 852
QY 3283 GGTGTACGAAGTGCACAGAAAGCTGTAAATGATATATGATATATAGATATATAGGATTTATGGAAA 3342
    |||||
Db 853 -----AspIleValGlyLysLeu--- 858
QY 3343 GAACAATGGAATATATATCAGATAAATACAAAGAAATTCATGACAGACCAAAATGCT 3402
    |||||
Db 858 ----- 858
QY 3403 GTTAGTAATAGTGTATGTAAAGCTTCAGACTGCCCCAAAATCATATAGACAGGAATGTT 3462
    |||||
Db 858 ----- 858
QY 3463 ATTGAATTTTGTGCGAATTTATACAACAAAATGTTGGCAAAAGTATATAAAGTGTACT 3522
    |||||
Db 858 ----- 858
QY 3523 AGTATGAAAGTGTCTGTCAATGTGTACTAACCCAGTATGAAATGTGTGAGCATATCTC 3582
    |||||
Db 858 ----- 858
QY 3583 CATGATACAGGAATTTTGTATGTGTCTCAGTCAAAAATGAGTTTGTGTATGAAAAAAGT 3642
    |||||
Db 859 -----PheGluAspAspLysSerLeuLysGlu--- 867
QY 3643 GATGTAGGATTAACGAAAAATATGCTTTAGAGATAAACACAGAGCATGATGTCG 3702
    |||||
Db 868 -----Ala 868
QY 3703 TGTGTTGTAAAGTGTGCGAAACGACAAAGGTTACAGTAAACGAAAAAAGGCG 3762
    |||||
Db 869 CysGlyLeuLysTyrGlyPro----- 875
QY 3763 GAAGAAAGATACGGAATTAACAGTAAATGATATACTTAAAGAAAAAGTGAAG 3822
    |||||
Db 876 -----GlyGlyLys 878
QY 3823 AACAAGATGAGATGTCATCCAAAAAGAAATAGTAATGATATGCCGATGCGCATGCG 3882
    |||||
Db 879 GluLys-----PheProAsnIlePheLysCys 886
QY 3883 GGAATATATAATTATGTC-----GAAGACCCCTGCTGTGTATGCC 3924
    |||||
Db 887 ValThrProSerGlyValSerThrIleThrSerGlyLysAspIleAlaIleCysValPro 906
QY 3925 CCTAGAGACAAAAGTTATGCGTACATTTCTTGCA-----AATGATAT 3969
    |||||
Db 907 ProArgIleArgIleLeuTyrValGlyGlyLeuSerGlnIlePheLysArgGlyLysAsp 926
QY 3970 GAATTAATAAATAATTAACATCAAGTT-----AAT 3999
    |||||
Db 927 GluThrThrGluValSerSerGluAlaThrSerAlaProSerGlnSerGlnSerGluLys 946
QY 4000 TTAAGAAAGACCTTCATCAATCTGCAGACAGCAAAACATTTCTTCATGATTTATATAT 4059
    |||||
Db 947 LeuArgThrAlaPheIleGlnSerAlaIleGlnThrPheLeuIlePheLysTyr 966
QY 4060 AAA-----AGTAAGGATGCT----- 4074
    |||||
Db 967 LysGlnGluLysLysProPheAlaThrGlnAspGlyAlaGlyLeuGlyValSerLeuPro 986
QY 4075 -----GAAGAAATGAACTCGATTAAGAAATTAAGAA-----GGCAAAATTCCT 4119
    |||||
Db 987 GluProSerProProGlyLysAspProGlnThrGlnLeuGlnIleThrGlyValIlePro 1006
QY 4120 CCCGATTTTGTAGATCATGTTCTACACATTTGGAGATTATAGCATTTTATTTTGA 4179
    |||||
Db 1007 ProAspPheLeuArgGlnMetPheTyrThrIleAlaIleAspTyrLysAspIleLeu----- 1024
QY 4180 ACAGATATATCAAAAGGTCTAGTGTGAGCAAGTAAACTAAAGACCAATAGATTTCTCT 4239
    |||||
Db 1024 ----- 1024
QY 4240 TTCAAAATGTGTACCAAAAATCTCTAATGCAAAAACAGCCAGATGTGACAGAA 4299
    |||||
Db 1024 ----- 1024
QY 4300 CATAGTCATGAGATATAGGAAGCTATGATGTGACTAGTAAATAATGGGCAAAAAA 4359
    |||||
Db 1024 ----- 1024
QY 4360 GATGATTTTACGAAAACTRAGGTTACACACAGTCGTAATTTAGTGACAAAGCACCACT 4419
    |||||
Db 1024 ----- 1024
QY 4420 TTGAGGAATTTGCCAAGACCCAGTTTTCAGATGGCTAACGGAATGTACGACGAC 4479
    |||||
Db 1024 ----- 1024
QY 4480 TATTGCTATACAGACAAAATATTGAAAGATGTGCGAAAAATGTAGTCAATGAC 4539
    |||||
Db 1024 ----- 1024
QY 4540 CAATTGACGTGTATACAGANTTAATTAAGAAATGCGAGAGCTAGCTTAAATATATGAAA 4599
    |||||
Db 1024 ----- 1024
QY 4600 AAAAAAAGATGATTCACAAAGATTAATTTACAGAGTGAACGCGCAAAAAAGAG 4659
    |||||
Db 1024 ----- 1024
QY 4660 TTCATAGACACACATTTGTGTATGTTTACAGACTATACGTAAGAAAGTACAGAGAT 4719
    |||||
Db 1025 -----TyrSerIleValIleLeuIleAspThrSer 1032
QY 4720 TACTGACAGAAATTTACTGCTAGTTGTGTGTATAGCTTGAAGTGCCTGTGTGTA 4779
    |||||
Db 1032 ----- 1032
QY 4780 CAAGAATATACATTTGTAGAAAAACAGCTTACTATGATGCCGACAAACATGTGGG 4839
    |||||
Db 1032 ----- 1032
QY 4840 TGCACAAAATTTATGAAAATGACGACAAATATATCTACATTTGAGTAAAGTAAAGTGC 4899
    |||||
Db 1033 -----AspThrThrGlyLysGlnThrProSerSerSerAsnAspAsnLeu 1047
QY 4900 AAAGATTTACTTAAAGGAGCAACACAGGTCTATTATAGTGCACAAAACAAAGTCTTAAT 4959
    |||||
Db 1048 LysAsnIleValIleLeuIleAlaSer----- 1055
QY 4960 AACTACAAATACCTGAAAGATTGACGAAAGATGTGCTTTTCCTCTCTGTCGACTACGT 5019
    |||||
Db 1055 ----- 1055
QY 5020 ATATGTTTCATGCAATGTGATGCAATTAATACAGATCCAGAAAGTTAAGATGAAAAATGGG 5079
    |||||
Db 1055 ----- 1055
QY 5080 TTGCGAAAAAGATGATGGAAGTGGCGGCAACGGAAGGTACAAATTTGGTCAATCTAC 5139
    |||||
Db 1056 -----GlySerThrGlu----- 1059
QY 5140 AAAGAAAAAAGAAAAAGAAAAATTAATAACCTCGAGATCCGACAAATATCTTATGAG 5199
    |||||
Db 1060 -----GlnIleLysGlnLysMetLys----- 1066
QY 5200 GTCCCGCTTGTAGTCTATGAATAATAGTTTATGATTTAAGAGATATATCTAGGT 5259
```



```
Db 1066 ----- 1066
OY 5260 ATTGATATTTGGAAGATGAAAAAACAAGACGAGAAAAATTGACAAATATTTAAC 5319
Db 1067 ----- 1077
OY 5320 AAAAAATGACATCAGTTGGCAAGAAGATAGTACTACAGGAAATCCCGTAGTACT 5379
Db 1078 ---GlyAlaThrSer-----GlyValProValThrLysAsnSerValLysThr 1093
OY 5380 GCGGAAATTTTCTGGAACGAAAAATAGAAATGTGTGGAAACGCAATGATATGCGGG 5439
Db 1094 ProGlnInhrTrpTrpGluAsnIleAlaLysAspIleTrpAsnIleMetValCysAla 1113
OY 5440 -----TACAACGTGTGTAGGATGATGAAATAGTGAATAGTGCACAGACATGATGA 5493
Db 1114 LeuThrLysLysGlu-----AsnAspAlaArg----- 1122
OY 5494 GATCTAAAAAATGTGGTGTGTACTCTCAGATGATGATTTCTTATGGGAAAAATTCGC 5553
Db 1122 ----- 1122
OY 5554 GATGAAGTACTGCGTATCAGTTTCTTCGATGGTTTCCGAAATGGGTGAGATTTTTCG 5613
Db 1123 ---GlyThrSer----- 1125
OY 5614 AAACATTAAGAAAAAGAAATTGAGAAATTCGTAGGGCGCTGAATGATTATCTGTGGT 5673
Db 1125 ----- 1125
OY 5674 GATAATGAAGTAAAAAGAAATGTACAGATGCGTGTACACAAATATAAAAATTTAAT 5733
Db 1125 ----- 1125
OY 5734 AGTAGAGTGAACACACAGATGAAAAAACAATAATATGTGAGATTAAGACAAA 5793
Db 1126 ----- 1133
OY 5794 ATATATTCGAGCATCTCTGTGGCAAAAGATGCAGAGACGCTGGCAATATTTAGACAAA 5853
Db 1133 ----- 1133
OY 5854 CAATTAATAAATTTGTGAATAATAAGTGGAGATTTGATATTAAGTATGAAAGAT 5913
Db 1134 ---LeuLysLys----- 1136
OY 5914 GTGTCCACACAGCATTAAGTATGATGATTAAGTCAAAATATGCGGCATATTAGACAT 5973
Db 1137 ----- 1147
OY 5974 GAACGAAAAAGAGTTGAAGAAAGTGTATTTGTCAAGTCCACAGAGTCCACACGTGTA 6033
Db 1147 ----- 1147
OY 6034 CGAAGGAAACACCGCTCACACAGGGTATCACTGATATCAAAAAGCAGCGCATCGAAAAA 6093
Db 1147 ----- 1147
OY 6094 GAAGGAAACAGCGCGCTTACAAAACAGCCGAAAAAAGTGGAAATCTAACAACAGAA 6153
Db 1147 ----- 1147
OY 6154 ATGCGACACAAACACGAACCCGACAGACAGACAAACAACAGAAACGAACATCAACA 6213
Db 1147 ----- 1147
OY 6214 GCACACACAAAGATCTGACGTGGCACAAATGTAAGGCCATCTTTGCAATAAACCA 6273
Db 1147 ----- 1147
OY 6274 GATAGCAGGGGTGGAATAGAGGTTTGAATCAAAAACGTATGACAAATATCTAAATGG 6333

Db 1147 ----- 1147
OY 6334 GGTGTATTTGTAGTAAGTCTAAAGAAAAATGAAAAATGCAATGATGATGCTCTAGAGAA 6393
Db 1147 ----- 1147
OY 6394 AAAAAATTAATTAATAATATACAAATTTTAATTAATTAATTAATTAATTAATTAATTA 6453
Db 1148 ----- 1159
OY 6454 AATGATTAATAAGAGCGCTTTTATTAATGTCAGCAATAGAACTCAATTTTGTGTTA 6513
Db 1159 ----- 1159
OY 6514 AATATATATTAATGAAAAATCTCCACACGAAAAATGAATTCGAAATTCAGAT 6573
Db 1159 ----- 1159
OY 6574 GAATTTAAAAAGAAATATGATATATACATATGATGATTAATAAGATATGTTTGGAACT 6633
Db 1159 ----- 1159
OY 6634 GATATTTCAATGATAAAAAATTAATTAACGTGAACAATAGTGAACAACCATTCGAA 6693
Db 1160 ----- 1160
OY 6694 GAAATTAATAGAAAAAACAGATTAATAAAGATTAAGAAATTAATTAATTTTGG 6753
Db 1161 GluSerGlyAlaLysSerAsnAsp----- 1168
OY 6754 GAGAAAAATTAATAATTTATTTGGGAAGAAATGATATATGATTAATTAATCATCTACA 6813
Db 1168 ----- 1168
OY 6814 GACGAAAAAGAAAAAATTAAGATTAATTAACAGTAACTGAATGACATGCCAACTG 6873
Db 1169 ----- 1171
OY 6874 ACGCTTCCCTTGAAGAGTTTGTAAAAAGGCCCAATTTTGAATGATGTTTACAGAAATGG 6933
Db 1172 ProProthrLeuLysAsnRheValGluLeuProthrPhePheArgTrpLeuHisGluTrp 1191
OY 6934 GCAGAGAAATTTTGTAAATAGAGAGAACACAGTTGTTAAATTTGAGCGCGGCTGT 6990
Db 1192 GlyAsnSerPheCysPheGluArgAlaLysArgLeuAlaGlnIleLysHisGluCysMet 1211
OY 6991 -----AAGAAATAGAGTGTAAAGTGTAAAGTGAAG----- 7026
Db 1212 AspGluAspGlyGluLysGlnTrp-----SerGlyAspGlyGluTrpCysGlu 1227
OY 7027 -----ACACAGAAATGGCAGAGCGG 7047
Db 1228 GluIlePheSerLysGlnTrpAsnValLeuGlnAspLeuSerSerCysAlaLysPro 1247
OY 7048 TGTGTAACATATCAAAATTTTATTAAGAGTGAAGAACTGAATATGAACACAAAGAA 7107
Db 1248 CysArgLeuTrpLysTrpIleGlnLysLysThrGlnLysGlnGlnLys 1267
OY 7108 AAGTTCAAAAAGATTAAGATGGAAGAAATATAGATATATCTCTACTGAAAGAGAC 7167
Db 1268 AlaTrpGlnGlnGlnLys-----SerAsnTrpLysGlnGlnGlnLysAsp 1282
OY 7168 ATAGAGAGCAACAGTGTCTCATGATATTTAAACATGAAATTAAGAATATATGTGC 7227
Db 1282 ----- 1282
OY 7228 AATAGGATTTGTCTGTATGCAAAAACCTTTTCAACACTACCAAAAACACACACAA 7287
Db 1283 -----LysCys-----GlnThr 1286
OY 7288 TCACAAATCATCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7341
Db 1287 GluSerAsnAsnAsnAlaSerGlnPheSerArgTrpLeuGlyAlaSerProThrAlaAla 1306
```

QY 7342 GAATTTAAACAAGTGTAGTCTCTGAACCTTTCAAAAAGGAGCTATGATTCATACAAAA 7401
|||
Db 1307 GluPhe-----LeuGlnIlySLeuIlySer----- 1314
QY 7402 AAAATTCGTAACCTAAATACCTATGATTCGTAGAGAAACAGCATATATTATTCCT 7461
|||
Db 1315 -----CysIlySAsnSpAsnIlyTyr----- 1321
QY 7462 AAAGAACGAAAAATATATGATATATACCTTGAGAAAAATTTATACCTATGAGTCT 7521
|||||
Db 1322 -----GluSngIlyGluAspAsnIlyIleAspPheIys---AsnProAspIysThr 1337
QY 7522 ACAAGAGAAAAAGAAAGTAAAAATAGTTGACATAATATATCTTCCTGCACTCCATAGAAA 7581
|||||
Db 1338 PheIysGluIleIlySer-----CysAspPro---Cys 1347
QY 7582 CCTATGACACCTGATTAATATATAGAGAAAGAAACCCCTGTGAAATATAGAGAAATAAT 7641
|||
Db 1348 ProIleThrGlyValIys-----CysGlnAsnGlyHis----- 1358
QY 7642 CGTTTAAAGTAGATATATGATGGAATGTACAAAAATTCAAAGTCTATCAGAGAAA 7701
1358
QY 7702 AAAAGAGTATGTACCTCCAGAGAGAACATATGTCTTAAGCAATTTAGATGAAT 7761
|||||
Db 1359 -----CysValGlySerAlaAsnGlyIlySgIlyCysIlySAsnAsn-----LysIle 1373
QY 7762 AAAATGGAAGACTTAAGATAGTAAATATATCTCTTAAAAATGCTTCGCAACCTGCACGA 7821
|||
Db 1374 ThrAlaGluAspIleIlySAsn-----LysThrAspPro 1384
QY 7822 AATGAGAGATAGACAT 7881
|||
Db 1385 AsnGlySngIleGluMetValIleSerAspAspSerThrAsn----- 1398
QY 7882 ATATGATATCTATGAAATATATAGTTTCTGCTGATCTGGGTGACATAGTTAGAGAACAGAT 7941
1399
QY 7942 ATGTTAGCAATGTGTGTTACTTACTTCCCTGAGCAATTAATATATATATATATATAT 8001
1406
QY 8002 TACATATGGAATAATGAGAAATTAATTAAGTAGAATAATATCAACAGATGTACAA 8061
1414 GlyIle----- 1415
QY 8062 ACGTTTGTCTGCTGTGGGATGCTAATAGAAAGATATTTGGAAA---GCAATGAGC 8118
1416
QY 8119 TGCAGAACACCAAGATGCAAAACTTTTAGAAAAAGAAAGATGATGCATTTTGAACGC 8178
1426 Cys-----GlyValAspIle 1430
QY 8179 ATACATTATATACAGAT---AAGTGTGACATTAAGAGACGATCCACCTGTTGATGATAT 8235
1431 CysThrIleuGluIlySngIlyIleSAsnGlyGlnGluIlyAsp-----LysIlyTyr 1447
QY 8236 ATACCTCAACGCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8295
1448 IleThrMetIys---GluLeuLeuIlyArgTrpLeuGluIlyThrPhe-----Leu 1462
QY 8296 GAAGCAATTTGAAAAATTTAAAAATTCATGTGATCAGTCT---AAAATCTGCAGAGATGC 8332
1463 GluAspIlySngIlyIleArgIlySngIlyIleSLeuSngIlySngIlySngIlySngIlySng 1482
QY 8353 AAGATATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8412
1483 Lys-----CysIleIlySngIlyCysIleGlu----- 1490

QY 8413 AATTTTCTTAAATGGAATCTCTATTTGATATACATCAAAATTAATTAACAAAGATTG 8472
1491
QY 8473 TATGAAACCAACATATATACAAAAATCTCTATGATCATGTTCAAAATTTTGTACAA 8532
1493
QY 8533 AAGTTGAAAACTTTTAAAGATGATGTTCTTGAGAGCTTTTCTGAAATCTTCATGAA 8592
1496 LysThrIlySngIlyTrpGln----- 1501
QY 8593 ACAAGTAAGTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8652
1502
QY 8653 TATGCTTGCAGAGAACCAACAAAAAGTATTAAGAAAGCTTGCACTGTACATCACTCTCT 8712
1507 Tyr-----LysIleAsnAsp-----Thr 1506
QY 8713 AAGATTCATTGATATATGCTTACAGATCAAAACAAAGATGAGATGAATTAACA 8772
1508
QY 8773 ACCTTTACCTCTGCTGCAAGAAATGATTAATTAATTAATTAATTAATTAATTAATTA 8832
1510 GlnTyr-----LysAsnAspAspGlyAsnThrIleuThrAsnPhe----- 1522
QY 8833 CTGTCTTAAATAGTTCAGATGATTAACAAAGGIGTATGATTCCTCCACAGAAAGACAT 8892
1522
QY 8893 TTATGTACAAGACCTATACATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 8952
1523
QY 8953 AAAAACTTCTTACTTCTGCTTTCAGTCAAGCAAAATTTAGGTCAAAAAATTAATTAATTCG 9012
1529
QY 9013 GAAGAGAGTGTGCTTTGAGGCAATGAATATATATATATATATATATATATATATATAT 9072
1530
QY 9073 AAAGAACTGATATGATGACACTTATATCTGAAAAAATTAATAAATAATTTGAACAA 9132
1537 LysProCysAspGlyLeuAsp----- 1543
QY 9133 TCAATGAGCAACCGAAAAATCGTAAAAACATGTGTGGAAAAATATATAGCGTCAAGATAG 9192
1543
QY 9193 CAGCGTATGTATGTGATTAATTAATTTCTACTTCAAAAGTAAACATTAATGATGAAGATG 9252
1543
QY 9253 TGTCATTTACCAAGATGAAGAAATAATACAGTTTCTGCTGTGTTAATTTGAATGGCA 9312
1544
QY 9313 AAGCAACGATGAAGAAAAAGAAACATGATGATTCATTAATAAACAAAAATGTCTCTGT 9372
1546
QY 9373 TCAACGAGATTAATTTTGAAGCTCAGAAATATATAGACAAACCTGATGTCAAGATGAT 9432
1552 AsnSerThrAspAsn-----SerGlnSngIlyAsnAsnAsnAsp 1564
QY 9433 ATTGAAAAATATATATACCTGATATATATGATTAATAAATAATCAATGGAATAATATA 9492
1565
QY 9493 AAATATTAAGCAATTAATAAGATCAATCTTCAGGTATATATAGACATAAACCATCTGAAGA 9552

```

Db 1576 LysIleSerGluCysLysGluGlnHisSerGlyLysIleThrGlnThr-----ProCysAsp--- 1593
OY 9553 AATGTGACATATATAAATCAAAAGATTCATGCCGCTTGGATTAATGATATA 9612
Db 1594 -----AsnSerSerLeuSerGlyLysGluSerThrLeuValGluAspValAspAspTyr 1611
OY 9613 AATGAATAGCTTACAGAACAAAAAATATGAATATAT----- 9651
Db 1612 GluGlu-----GlnAsnProGluAsnLysValGluGlnProLysPheCys 1626
OY 9652 ---GAATTCAAAGAGACTATAAATAATTTATTCCTGCTTATATTTGTTGAAGATGA 9708
Db 1627 ProAspSerLysGluProLysLysGluAsnAspGluValGlyThrCysGlyLysP 1646
OY 9709 ACACACAAAAATCATGACTAGATGAAATATA-----AAAGAGAGAGCAACAGT 9762
Db 1647 GluGluLysLysLysValGluAspSerValIleGluGlnLysGluGluAlaIleAs 1666
OY 9763 GGTCCCTAAAGCACTATTTCTTACACCCCATGTAGATTTCTTATCAAGCAGCTTAA 9822
Db 1667 AlaProGluLysSerProProLeuThrPro-----GluAlaProLys 1680
OY 9823 TTCTCAACATGAGTACAGACATAATGATCT----- 9855
Db 1681 LysGluGluAsnValValProLysProProProProLysLysArgIleLysThr 1700
OY 9856 AAAAAAT-----GATATTTGAAAAAGTATATCTGTTGTTATGTTATG 9900
Db 1701 ArgAsnValLeuAspHisProAlaValIleProAlaLeuMetSerThrIleMetTyr 1720
OY 9901 GCCTTACGTTTG-----ATAGCGCTTACATTCATGAGAAAAATTCAAATCG 9948
Db 1721 SerIleGlyIleGlyPheAlaAlaPheThrTyrPheTyrLeuLysLysThrLysSer 1740
OY 9949 TCTGTG---GACTTGTGGTATACGATACGATATCCCGCAAGAGACATGATGATGCTTAC 10005
Db 1741 SerValGlyAsnLeuPheGlnIleLeuGlnIleProLysSerAspTyrAspIleProThr 1760
OY 10006 TTGGAAATCCAAAAATAGTACATACATAGAAAGTGTCCATATTAAGCAAAACATAT 10065
Db 1761 LeuLysSerSerAsnArgTyrIleProTyrAlaSerAspArgHisLysGlyLysThrTyr 1780
OY 10066 ATATATATGAGAGAGATCTGAGATGAGATGAGATTAATATATGCGACTTATCTTCC 10125
Db 1781 IleTyrMetGlnGlyAspSerSerGlyAspGlu---LysTyrAlaPheMetSerAspThr 1799
OY 10126 TCTGATATTACTTATCCGAAGTGAATGATGAAATGATTAATGATATATATGTA 10185
Db 1800 ThrAspIleThrSerSerGluSerGluTyrGluGluLeuAspIleAsnAspIleTyrVal 1819
OY 10186 CCAGGTAGTCCCTAATATATAAATGATGATGAGATGATGAAACCATCAAAAAAGGAT 10245
Db 1820 ProGlySerProLysTyrLysThrLeuIleGluValIleGluProSerLysArgAsp 1839
OY 10246 ATACCAAGGAT-----GATTACACCAAGTAATGATACACACAGT 10294
Db 1840 ThrGlnAsnAspIleHisAsnAspIleProSerAspIleProSerAspThrProPro 1859
OY 10285 ACGAATAGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10344
Db 1860 Pro-----IleThrAspAspIleTyrPheAsnGlnLeuLysLysAspPheIleSerAsnMet 1877
OY 10345 TTA-----CCAAATACAGAACCAAAATATAATTAATAAAGTGCAGATATTCCAATG 10395
Db 1878 LeuGlnAsnThrGlnAsnThrGluProAsnIleLeuHisAspAsnValAsp-----Asn 1895
OY 10396 AATACAGAACCAAAATATATATATATATATATATATATATATATATATATATATAT 10455
Db 1896 AsnThrHisPro---ThreSerArgHisAsnMetAspGlnLysPropheIleMetSer 1914
OY 10456 ATTCAATGATAGGATTTATATATATGAGAAAGAAATTAATGAT----- 10497

```

```

Db 1915 IleHisAspArgAsnLeuPheSerGlyGluGluTyrAsnTyrAspPheAsnSerGly 1934
OY 10498 -----AATATTAATAGTACTAAT-----ACTAATAAT 10527
Db 1935 AsnAsnProIleAsnIleSerAspSerThrAsnSerMetAspSerLeuThrSerAsn 1954
OY 10528 GATATCCAAATGATATCTGAAATGATTTCTTATAGAGGATATGATTAATTAATGATCA 10587
Db 1955 HisSerProTyrAsnAspLysAsnAspLeuTyrSerGlyIleAspLeuIleAsnAspAla 1974
OY 10588 CTAGTGTCTTAAACCTATATATATATATATGATGATGATG 10628
Db 1975 Leu---SerGlyAsnHisIleAspIleTyrAspGluMetLeu 1987

```

RESULT 10

```

AAV77906
ID AAV77906 standard; Protein: 2182 AA.
XX
XX AAV77906;
AC
XX
XX 13-JUN-2000 (first entry)
DT
XX
XX Plasmodium var-1 polypeptide.
DE
XX
XX DBL gene; Duffy-Binding Like gene; ebl-1; Duffy Antigen Binding Protein;
KW DABP; Static Acid Binding; Protein; SAMP; malaria; vaccine; immunisation;
KV protozoacide; var-1.
XX
XX Plasmodium sp.
OS
XX
XX US593827-A.
PN
XX
XX 30-NOV-1999.
PD
XX
XX 07-JUN-1995; 95US-0487826.
PF
XX
XX 10-SEP-1993; 93US-0119677.
PR
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA
XX
XX Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;
PI
XX
XX WPI: 2000-194198/17.
DR
XX
XX N-PSDB; AA298288.
DQ
XX
XX Isolated protein binding domains from Plasmodium vivax and Plasmodium
PT falciparum erythrocyte binding proteins useful for vaccinating against
PI malaria.
XX
XX Disclosure; Columns 129-140; 93pp; English.
XX
XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
XX (Duffy-binding like) gene family. The ebl-1 proteins are substantially
XX identical to the Duffy Antigen Binding Protein (DABP) and Static Acid
XX Binding Protein (SABP), which are soluble proteins that appear in the
XX culture supernatant after erythrocytes infected with malaria release
XX merozoites. Immunochemical studies indicate that DABP and SABP are the
XX respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
XX and static acid receptors on erythrocytes. The ebl-1 polypeptides may be
XX used to vaccinate against malaria, especially caused by P. falciparum.
XX Immunization with the polypeptide provides effective protection against
XX malaria. The present sequence represents the var-1 polypeptide.
XX
XX Sequence 2182 AA.
SQ

```

Alignment Scores:

```

Pred. NO.: 5.54e-115
Score: 1755.00
Percent Similarity: 29.37%
Best Local Similarity: 20.35%
Query Match: 9.02%
DB: 21
Gaps: 125

```


QY 2005 GAAGGTATTTTAAAGTTATGATTAACCTTGACAAAGATGACAAATGAGAAAGAA 2064
Db 691 ValPheSerProTyrGlyValLeuAspLeuVal-----LeuYsGlyGlyAsn 706
QY 2065 CTATATGSAATATATATAAATAAATAAATGACGTTTCCATTGGAATAATATAGGAC 2124
Db 707 LeuLeuGlnAsnIleYsAspValHisGlyAspThrAspIleYsHisIleYsYs 726
QY 2125 TATTTAGAG-----AATGCAATA 2142
Db 727 LeuLeuAspGluGluAspAlaValAlaValLeuGlyGlyYsAspAsnThrIle 746
QY 2143 GAACCTCTGTATGATCACTTAAAGAACTGCCAGATATGTAAGACATAATACAAAC 2202
Db 747 AspLysLeuLeuGlnHisGlyGluGluGlnAlaGluGlnCysIleYsGlnYs-----Gln 764
QY 2203 GAAGCACTGTGAACATCCCATATGCAACAACAACCCGCTGTAAACCTCTGTGAGGC 2262
Db 765 GluGluCysGlu----- 768
QY 2263 ACGCAACCCACTTAAATAATATAAAGAAATAGCACAACTACTTTAAAGAGATGATACGAG 2322
Db 769 -----LysIleAlaGlnGln 773
QY 2323 GAAGCAGAAATCGGTGCTTCTATAATTGAAAGAAAGGACACAGAGATATATATAA 2382
Db 774 GluSerArgGlyIleYsSerAlaGlu----- 781
QY 2383 CGTGGGGTAGAGAAAGCACTTCAAGACAAATTTATGATGATATATGATTAACATTC 2442
Db 782 -----ThrArgLysGluIleYs----- 787
QY 2443 AATCGTAATCTGGTTTTCAAATGACCATGTGATGCGAAAGGACAGATGATGTATA 2502
Db 788 -----ThrGlnGlnProAlaAspSerAlaGlyGlu----- 797
QY 2503 CAACAAGATTTGTGTAGAACTGATGGAAGTGGATCCGGAACACATGCGTAAAGAT 2562
Db 798 -----ValGluGluGluGluAspAspAspIleYsAspGlu 810
QY 2563 CACGAGATGTTATATATGCTCCTAGAAAGACACATATGTACATCCAAATTTGAAACAT 2622
Db 811 AspGluAsp----- 813
QY 2623 TTACAAACGATGATCACCCACTTAATGCTAATATGTTGATGATTTAGTTAATATCC 2682
Db 814 -----AspAspValValGlnGlnGlu 820
QY 2683 TTTTGGGGATGTTCTCTATCAGCAAAATATGAAAGCAACAAGATATATCGATGAT 2742
Db 821 GluGluGly----- 823
QY 2743 AAAGAAAGATTAACCTTAAAGGGCCCAAGAGATGACGCCAAACACAGACAAT 2802
Db 824 LysGluGluGluGlyThrValThrGluValThrGluValThrGluValGluGluIleYs 842
QY 2803 ATCTGTGAGCATATCGTTACGTTTGCAGATATAGGTATATATTCGAGGAAGAGAT 2862
Db 842 ----- 842
QY 2863 CTCTGGAAAGAAAGGTGACATGTAAGCTGCAAGACATTTGGAACGTTTGTGT 2922
Db 842 ----- 842
QY 2923 AATATACATAAGTCACTCAAGAGCAAGAAAGATATATATATGATGATGCCCCCAA 2982
Db 842 ----- 842
QY 2983 TATTTAAATTTGAGGAAATTTGGTGGAGACTAATAGACCAAGTATGGAAGCCATG 3042
Db 842 ----- 842
QY 3043 AAATGTGATATAAATATTTGAAGATTAATCGGAGACCAATCAACAAAGTATAT 3102

Db 842 ----- 842
QY 3103 TCGGATATATGATCATATACACCATTTGATGATATATATCCCAAAATAATAGATGATG 3162
Db 843 -----Val 843
QY 3163 ACCGAATGGCGAATGTGACTGCAAGGTGCGAAGAAAGAGTATGATTAAGTGAAGAG 3222
Db 844 ThrGlu-----GlnGlu 847
QY 3223 AAGTGAAGAGTGAAGATTAAGATTAAGATTAAGTCAAGGCTGTACGAAGAGTGTACA 3282
Db 848 GlyValIleYsProCys----- 852
QY 3283 GGTGTACAGAGTGCACAGAGCTTGTATGATATATATATATATATATATATATGAAA 3342
Db 853 -----AspIleValGlyLeu----- 858
QY 3343 GAACAATGGAATCT 3402
Db 858 ----- 858
QY 3403 GTTAT 3462
Db 858 ----- 858
QY 3463 ATTGAATTTTGTGCGAATTTATCCACAAATAATGCTGCAAAAGTATATATATATATAT 3522
Db 858 ----- 858
QY 3523 AGTGATGAAGTGTGCTATGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3582
Db 858 ----- 858
QY 3583 CATGATACAGAAATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3642
Db 859 -----PheGluAspAspIleYsSerLeuYsGlu----- 867
QY 3643 GATGTAAGATTAACGAAATAATATATATATATATATATATATATATATATATATATAT 3702
Db 868 -----Ala 868
QY 3703 TGTGTTGTAATAAGTGTGTCGAAACGACAGAGGTATGATTAATAAAGAAAGAAAGCG 3762
Db 869 CysGlyLeuYsTyrGlyPro----- 875
QY 3763 GAAGAAAGATACGAAATGTAATAACAGTGAATATATATATATATATATATATATATAT 3822
Db 876 -----GlyGlyYs 878
QY 3823 AAACAAGTACAGATGTCATCCAAATAAGATGATATATATATATATATATATATATATAT 3882
Db 879 GluYs-----PheProAsnThrPlyCys 886
QY 3883 GGAAT 3924
Db 887 ValThrProSerGlyValSerThrAlaThrSerGlyYsAspGlyAlaIleCysValPro 906
QY 3925 CCTGGAAGACAAAGTATATGCTGATATTTCTTGCA-----AATGATAT 3969
Db 907 ProArgIleArgIleArgLeuTyrValGlyLeuSerGlnIlePheSerAlaArgGlyAsp 926
QY 3970 GAATATAAATAAT 3999
Db 927 GluThrThrGluValSerSerGluAlaThrSerAlaProSerGlnSerGluSerGluYs 946
QY 4000 TTAAGAAAGCTTTCATCAATCTGCAGCAGCAAGAAACATTTCTCATAGTATATATAT 4059
Db 947 LeuArgThrAlaPheIleGluSerIleAlaIleGluThrPhePheLeuThrPheIleYsTyr 966
QY 4060 AAA-----AGTAAAGATGCT----- 4074

```
Db 967 LysGluLulysLysProProAlaThrGlnAspGlyAlaGlyLeuGlyValSerLeuPro 986
QY 4075 -----GAGGAATTAAGCTCGATTAAGAAATTAAGAA--GGCAAAATTCCT 4119
Db 987 GluProSerProProGlyGluAspProGlnThrGlnLeuGlnGlnThrGlyValIlePro 1006
QY 4120 CCGCATTTTGGATGCCATGCTCTACATTTGGAGATTATAGAGTTTATTATTGGA 4179
Db 1007 ProAspPheLeuAlaTrgGlnMetPheYrThrLeuAlaAspYrLysAspIleLeu 1024
QY 4180 ACAGATATATCAAAAGGTCATGCTGAGGAGAGTAACATAAAGCAAAATAGATTCTCT 4239
Db 1024 ----- 1024
QY 4240 TTCAAAATGTGACCAAAAATCTCCTAATGAAAAACAGCCAGAATGTGACAGAA 4299
Db 1024 ----- 1024
QY 4300 CATAGTCATGAGATATGGGAAGCTATGCTATGTGCACATGTAATAAATTGGGGCAAAAA 4359
Db 1024 ----- 1024
QY 4360 GATGATTTTACGAAACCTACGGTTACACACAGTCAAAATTTAGTACAAAAGCACACT 4419
Db 1024 ----- 1024
QY 4420 TTGGAGAAATTTGCCAAACGACCCAGTTTTCAGATGGCTAACGAATGTACAGAC 4479
Db 1024 ----- 1024
QY 4480 TATTGCTATACAGCAAAAATATTGGAAGATGTGACAGAAAATGTAAATGATGAC 4539
Db 1024 ----- 1024
QY 4540 CAATTGAAGTGTATACAGATTAATAAGAAATGCGAGACTACGTTAAATATATGAAA 4599
Db 1024 ----- 1024
QY 4600 AAAAAAAGATGTGATCCACAGATTAATATTACAGAGTGAACGACGCAAAAAAGA 4659
Db 1024 ----- 1024
QY 4660 TTGATAGACACACATTTGTTAATGTTACAGACTATCTGGAAGCAATGCAACGAT 4719
Db 1025 -----TySerGlySerAsnAspThrSer 1032
QY 4720 TACTTGAACAGAAATTTACTGCTAGTTGTGTGATTAAGCTGGAAGTGCCTGTGTGA 4779
Db 1032 ----- 1032
QY 4780 CAAGAATAATACAAATTTGAGAAAAACAGCTTACTATGATGCCGACAAACATTTGGG 4839
Db 1032 ----- 1032
QY 4840 TGCACAAAATTTATGAAAAATGACGACAAATATACATACTTTCGAGTAAAGTAAAGTGC 4899
Db 1033 -----AspThrThrGlyLysGlnThrProSerSerAsnAspAsnLeu 1047
QY 4900 AAAAGATTACTAAAGGCAACACAGAGTCTATTAAAGTGCAAAACAAAGGCTCAAT 4959
Db 1048 LysAsnIleValIleuGlnAlaSer 1055
QY 4960 AACTACAAATTAAGTGAAGAATTAAGTGAAGATGTGCTTTTCTCTCTCGTCACTACGT 5019
Db 1055 ----- 1055
QY 5020 ATATGTTTTCATGATGATGATGGCAATTAACAGATCCAGAAAGTTAAAGATGAAGATGG 5079
Db 1055 ----- 1055
QY 5080 TTGCAAAAAAGATTGATGAAGTGGCGCAACGAAAGGTACAAATTTGGCTCAATACTAC 5139
Db 1056 -----GlySerThrGlu 1059
QY 5140 AAAAAAAGAAAAAGAAAAATTAATAAAGCTCGAGTCCGACAAATATCTTATGAG 5199
Db 1060 -----GlnLulysGlnLysMetLys 1066
QY 5200 GTCCCGCCTTGATGCTATGAATATAGTTTTATGATTTAAGATATATATTCTAGGT 5259
Db 1066 ----- 1066
QY 5260 ATTGATTAATTTGGAATGAAAAAACAAGACGACGAAAAATTTGAGAATAATTTAAC 5319
Db 1067 -----GlnIleGlnAlaLysIleLysIleLeuAsn 1077
QY 5320 AAAATGGAACATCAGTTGGCAAGAAGATGATAGTACTACAGAAATCCCGTACTACT 5379
Db 1078 ---GlyAlaThrSer-----GlyValProValThrLysAsnSerValLysThr 1093
QY 5380 GCGCAAAATTTTCTGGAACGAAAAATTAAGATGTGTGGAACCATGATATCGGG 5439
Db 1094 ProGlnGlnThrPyrGluAsnIleAlaLysAspIleTrpAsnIleMetValCysAla 1113
QY 5440 -----TACAAAGTGTAGGATGATGAAATAGTGAATATGTGCAAGAACTGATGAA 5493
Db 1114 LeuThrTyrLysGlu-----AsnAspAlaArg----- 1122
QY 5494 GATCTAAAAAATGTGGTCTGTACCTCAGATGATGATATATCTATGGGAAAAATCCG 5553
Db 1122 ----- 1122
QY 5554 GATGAAGTACTGCGTATCAGTTTCTTCGATGGTTGGCAATGGGGTGAAGATTTTGC 5613
Db 1123 -----GlyThrSer 1125
QY 5614 AAACATAAAGAAAGAAATTTGAGAAATTTGTAGGGCGTGAATGATTTACTTGGGT 5673
Db 1125 ----- 1125
QY 5674 GATTAATGAAGATAAAGAAAGAAATGTACAGATGCGTGTACAAATATAAATAATTAAT 5733
Db 1125 ----- 1125
QY 5734 AGTGAATGAAACACAGTATGAAAAAACAATAATATGCTGAATTAAGACAAA 5793
Db 1126 -----AlaLysIleGlnLulysAsnPro 1133
QY 5794 ATATATTCGAGATCCTCGTGCAAAAGATGACAGAGAGCTGCGCAATATTAGACAAA 5853
Db 1133 ----- 1133
QY 5854 CAATTAAAAAATTTGTGAAAAATTAAGTGAAGATTGGAATATAAGTATGAAAGAT 5913
Db 1134 ---LeuLysLys----- 1136
QY 5914 GTGTCCACACAGGATTAATGATGTAATAGTCAAAATATGCCCCGATCATTAAGCAGAT 5973
Db 1137 -----AlaLeuTrpAspGluAlaAspLysAsnThrPro 1147
QY 5974 GAACCAAAAGAAAGTGAAGAAAGTGAATTTGCAAGTGCACAGAGTCCACACAGTGA 6033
Db 1147 ----- 1147
QY 6034 CGAAGGAAACACCGTCCACAGGGTATCACTGATATCAAAAGCAGAGGATGAAAAAA 6093
Db 1147 ----- 1147
QY 6094 GAAGCAAAACAGCGCCGCTACAAAAACAGCCGAAAAAGTGAATAATCTAACACAGAA 6153
Db 1147 ----- 1147
QY 6154 ATGCGACACAAACAGAACCCGACGACGACACAAACAGAAACGAAACGAAATCAACA 6213
Db 1147 ----- 1147
```



```
OY 6214 GCACACACAGCAATCTGACGTGGGCACAAATGTAAGGCCATTCTTTCAGATTAACCA 6273
Db 1147 -----
OY 6274 GATACGAGGGGTGGAATAGAGGGTTGTATCCAAAAACGTATGACATATCTTAATGG 6333
Db 1147 -----
OY 6334 GGTGTATTTAGTAACTTAAGAAAAATGAAATGGCATATGTATGCTCTAGGAGA 6393
Db 1147 -----
OY 6394 AAAAAATATGTATTAATATATCAATATTTAATTATGAACCTGAATAATAGCGTGAC 6453
Db 1148 -----
OY 6454 AATGATATAAAGAGCGTTTATTAATATGCGACCAATAGAACTCAATTTTGTGTGA 6513
Db 1159 -----
OY 6514 AATATATTAATTTGAAATCTCGACAGAAAAATGATTCGAAAAATGGACATTCAGAT 6573
Db 1159 -----
OY 6574 GAATTTAAAGAAATATGCTATTTATACATATGTGATTAAGATATGTTTTTGGAACT 6633
Db 1159 -----
OY 6634 GATATTTCTAATGATTAATAAATAATTAATCTGAACAATAGGTACAAACATCTCAT 6693
Db 1160 -----
OY 6694 GAAATTAATTAAGAAAAACGGTAAAAAAGATGAAGAAATTAAGTAAATATTTTGG 6753
Db 1161 -----
OY 6754 GAGAAAAATTAATTTATTTGGAAAGATGATATGATTAATCTATCATCTCACA 6813
Db 1168 -----
OY 6814 GACGAAACGAAAAAAGAAAAATTAAGATTAATTAACAGTACATGACACCAACTG 6873
Db 1169 -----
OY 6874 ACAGCTCCCTGGAAGATTGTAAGAAAGCCCAATTTTGAGATGCTTCACAGAAATGG 6933
Db 1172 -----
OY 6934 GCAGAGAAATTTTGTAAATAGAGAGAAAGCAAGTGTAAATTTGAGCGCGCGCTG 6990
Db 1192 -----
OY 6991 -----
Db 1212 -----
OY 7027 -----
Db 1228 -----
OY 7048 TGTGTAAACATATCAAAATTTTATTAAGAACTGAAACTGATATGAAGACAAAGAA 7107
Db 1248 -----
OY 7108 AAGTCAAAAGATTAAGATGCGCAAAAGATTAAGATTTCTCTTACTGAAAGAGAC 7167
Db 1268 -----
OY 7168 ATAGAGAGGACATGCTCATGATTAATTAACATGAATTAAGAAATTAATGCGC 7227
Db 1282 -----
OY 7228 AATAAGATTTGTTCTTGTATGCAAAACCTTCTTCAACAATACCAAAAAACACAA 7287
Db 1282 -----

OY 1283 -----
Db 1283 -----
OY 7288 TCACATCATCGAGATGCTAATGATATGCCAATGCTGATTAATGTTCT 7341
Db 1287 -----
OY 7342 GATTTTAACAGTGTAGTGTCTGACCTTCAAAAAAGCGATATGATTCATCAAAA 7401
Db 1307 -----
OY 7402 AATATCTGACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7461
Db 1315 -----
OY 7462 AAGAGACGAAATTAATGATATTAATTAATTAATTAATTAATTAATTAATTAAT 7521
Db 1322 -----
OY 7522 ACAAAGAAAGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7581
Db 1338 -----
OY 7582 CCTATGACCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7641
Db 1348 -----
OY 7642 CGTTTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7701
Db 1358 -----
OY 7702 AAAAGATAGTGTACCTCCAGAGAGACATGCTTGAAGAAATTAATTAATTAAT 7761
Db 1359 -----
OY 7762 AAAATGAAAGCTTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7821
Db 1374 -----
OY 7822 AATGAGAAATGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7881
Db 1385 -----
OY 7882 ATATGATATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7941
Db 1399 -----
OY 7942 ATGTACCAATGCTGCTTACTTACCTCCGTAAGAAATTAATTAATTAATTAAT 8001
Db 1406 -----
OY 8002 TACATATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8061
Db 1414 -----
OY 8062 ACCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8118
Db 1416 -----
OY 8119 TCGAAAGACACAGACATGCAAAATTTTAAAGAAAGAAATGATGATTAAGAGC 8178
Db 1426 -----
OY 8179 ATTAACATTAATCAAGAT 8235
Db 1431 -----
OY 8236 ATTCCTCAACGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 8295
Db 1448 -----
OY 8296 GAGAAATGGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8352
```

Db 1463 GLUASPTYASNAIRgIleatgIlyslIlelyLeucThrlYslsgIuaspGlyCys 1482
QY 8353 AAGAAATGATTTATGATGATAAATAAGTGAACAGTGTAAAAAGAGATGCAAGATATATAA 8412
Db 1483 Lys-----CysIlelyslGlyCysIleGlu----- 1490
QY 8413 AATTTGTCTTAAATGGAATCTCTATTCGATATACATCAATAATATACAAAGAATTG 8472
Db 1491 -----LysTrp----- 1492
QY 8473 TATGAACAACCAATATATACAAAATCTCTACTTATGATCATGTGCAAAATTTTGTACA 8532
Db 1493 -----ValGln-----Glu 1495
QY 8533 AAGTGAACCTTTTAAAGTGAATGTCTGTGAGAGCTTTTCTGAATATCTTCATGAA 8592
Db 1496 LysThrIlysgIuIurpGln----- 1501
QY 8593 ACAAGTAAAGTGTGATTTAAATTTAATGAAAATGANGTCTCTCCAAATATACGACA 8652
Db 1502 -----LysIleasnAsp-----Thr 1506
QY 8653 TATGCTTTCGAAGAAACACCAAAAAGTTATAAAGAGCTTGACAGTTGTACACTACTCT 8712
Db 1507 Tyr----- 1507
QY 8713 AAGAAATCATGATGATATATGCTCTACCGATCAAAACAAGATGATGATGATGATGATGAT 8772
Db 1508 -----LeuGlu 1509
QY 8773 ACTTTTACCTTCTGTCGGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 8832
Db 1510 GlnTyr-----LysasnAspGlyasnThrLeuThrAsnHe----- 1522
QY 8833 CTGTCTTAATAGTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8892
Db 1522 ----- 1522
QY 8893 TTTATGACAGACCTATACATGATATATATGATGATGATGATGATGATGATGATGATGAT 8952
Db 1523 -----LeuGlnIlePheGlnTyrArg----- 1529
QY 8953 AAAAACTTCTACTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT 9012
Db 1529 ----- 1529
QY 9013 GAAGAAGACTGTGCTTTCAGCAATGAATATAGTTATGACATTTATTCGATATATAT 9072
Db 1530 -----ThrGluPheIlysnAlaIle 1536
QY 9073 AAGAAGACTGATATGATGACACATCTATATATGATGATGATGATGATGATGATGATGATGAT 9132
Db 1537 LysProCysaspIlyLeuasp----- 1543
QY 9133 TCAATATACCAACCCAAAATCGTAAACATGTGTGGAAAATATATGAGCTCAGATATGG 9192
Db 1543 ----- 1543
QY 9193 CACGCTATGTTATGTGATATATAAATGCTACTTCAAAAGTACATATGATGATGATGATG 9252
Db 1543 ----- 1543
QY 9253 TGTCAATTTACCAAGATGAGAGAAATCACTTTCTGCTGTTAATTGATGATGATGATG 9312
Db 1544 -----GlnPhe----- 1545
QY 9313 AAGCAAGATGATGAGAGAGAAATCATGTATGATGATGATGATGATGATGATGATGATGAT 9372
Db 1546 -----LysThrSerCysGlyLeu 1551
QY 9373 TCAACGAAGATATATTTTGAAGCGTCAGAAATTTATGACACACTGATGATGATGATGATG 9432
Db 1552 AsnSerThrIlysnAsp-----SerGlnsnGlyasnAsnAsp 1564

QY 9433 ATTAGAAAATATATATAGCTTGAATATATGATATATATATATATATATATATATATATAT 9492
Db 1565 -----LeuValLeucCysLeuLeuasnIlyslGlnIly 1575
QY 9493 AATATATGCAATTTAAAGATCAATCTTCAGTATATATGACATATATATATATATATATAT 9552
Db 1576 LysIleSerGlnCysIlysgIuGlnIleSerGlyGlnThrIleThr-----ProCysAsp 1593
QY 9553 AATGTACAGAT 9612
Db 1594 -----AsnSerSerLeuSerGlyIlysgIuSerThrLeuValGlnIlysnPvalAspAspTyr 1611
QY 9613 AATGAATATGTTACAGAACCAAAAATATGAAAATAT----- 9651
Db 1612 GlnGlu-----GlnasnProGlnIlysnIlysnValGlnIleProIlysnPheCys 1626
QY 9652 ---GAATTCAGAGACTACTATAAAAATATATATATCTGCTTATATATATATATATATAT 9708
Db 1627 ProAspMetIlysgIuProIlysnIlysgIuasnAspGlnIlysnIlysnIlysnIlysnIly 1646
QY 9709 ACACAGAAAAT 9762
Db 1647 GlnGlnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysn 1666
QY 9763 CGTCTTAAGCACTCTATTTCTTTACACCCCATGATGATGATGATGATGATGATGATGATG 9822
Db 1667 AlaProGlnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIly 1680
QY 9823 TTTCTAACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9855
Db 1681 LysGlnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIly 1700
QY 9856 AAAAAAT-----GATATATGAAAAGATGATGATGATGATGATGATGATGATGATGATG 9900
Db 1701 ArgasnValLeuasnPheIleProAlaValIleProIleLeuLeuSerSerThrIleMetTrp 1720
QY 9901 GCCTTATGTTT-----ATAGCGCTTCATTTGATGATGATGATGATGATGATGATGATGAT 9948
Db 1721 SerIleGlyIleIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIly 1740
QY 9949 TCTGTG---GACTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10005
Db 1741 SerValGlyasnLeuPheGlnIleGlnIleProIlysnIlysnIlysnIlysnIlysnIly 1760
QY 10006 TTGCAATCCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10065
Db 1761 LeuIlysnSerSerasnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIly 1780
QY 10066 ATATATATGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10125
Db 1781 IleTyrMetGlnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIly 1799
QY 10126 TCTGATATTTACTTATCCGAAAGTGAATGATGATGATGATGATGATGATGATGATGATGAT 10185
Db 1800 ThrAspIleThrSerSerGlnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIly 1819
QY 10186 CCAGTAGTCTTAAT 10245
Db 1820 ProGlnSerProIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIly 1839
QY 10246 ATACCAAGTAT-----GATACACCAAGTATGATGATGATGATGATGATGATGATGAT 10284
Db 1840 ThrGlnasnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIlysnIly 1859
QY 10285 ACGAATATGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10344
Db 1860 Pro-----IleThrAspIlysnIlysnIlysnIlysnIlysnIlysnIlysnIly 1877
QY 10345 TTA-----CCAATATACAGAACCAATATATATATATATATATATATATATATATAT 10395
Db 1878 LeuGlnasnThrGlnasnThrGlnProasnIleLeuHisAspasnValAsp-----Asn 1895

OY	10396	AATACGAGACCCTAATCTTATTTCTGTGAATTCGGAAAGAAAACCTTTATTANACT	10453
.Db	1896	AsnThrHisPro---ThrMetSerAlaGlnIleAsnMetaspCInlysrPheIleMetSer	1914
OY	10456	ATTCAATGATNAGGATTTATTAATACTGGGAAAGAAATAGTTAT	10497
.Db	1915	IleHIsAsPArqAsnLeuHeserGylgluGlutYAAsntYAAspmethHeasnserGyl	1934
OY	10498	-----AATTTAATATNAGTACTAAT-----ACTAATAT	10527
.Db	1935	AsnAsnProIleasnlIeserAspsScrrTrhnSernMetaspSerLeuthrSerAsn	1954
OY	10528	GATATTCCAATGAAATGCTAGAATGATTTCTTAGAGTATAGATTATTAATGATCA	10587
.Db	1955	HISserPROtYAAsnspysAsnspleutyrsereGyllleAspleuileasnspsala	1974
OY	10588	CATAGTCGTCTTAACCATTTGATATATANGATGAAGACTATG	10628
.Db	1975	Leu--SerGlyAsnHIsleAsplletyryAspGlumetleu	1987
<hr/>			
XX	JULF 11		
XX	93944		
XX	AAW93944 standard; protein; 2228 AA.		
XX	AAW93944;		
AC			
XX			
DT	30-JUN-1999 (first entry)		
XX			
DE	P. falciparum PfEMP1 protein.		
KM	Erythrocyte membrane protein; EMP, PfEMP1, malaria; antioclusional;		
KM	glycosaminoglycan-like moiety; antiaggregational; antimalaria;		
KW	antigen receptor; infected erythrocyte; rosette formation; blood cell;		
KW	capillary occlusion; cerebral malaria; treatment; vaccine; detection;		
KW	medicament; parasite; diagnosis; drug screening.		
XX			
OS	Plasmodium falciparum.		
XX			
PN	WO991557-A1.		
XX			
PD	01-APR-1999.		
XX			
PF	18-SEP-1998; 98WO-SER01675.		
XX			
PR	19-SEP-1997; 97SE-0003386.		
PA	(KARO-) KAROLINSKA INNOVATIONS AB.		
XX			
Barragan A, Carlson J, Fernandez V, Qijun C, Wahlgren M;			
WPt; 1999-254692/21.			
XX			
PT	New isolated malaria polypeptides		
XX			
PS	Claim 4; Page 67-74; 80pp; English.		
XX			
CC	This invention describes a novel Plasmodium falciparum erythrocyte		
CC	membrane protein (EMP), PfEMP1, which is capable of binding to a		
CC	carbohydrate which exhibits at least one negatively charged		
CC	glycosamino-glycan (GAG)-like moiety and has antiaggregational,		
CC	anticoclusional and antimalarial activity. The carbohydrates of the		
CC	invention are capable of acting as receptors for malaria antigens		
CC	present on the surfaces of malaria infected erythrocytes, by binding		
CC	to these antigens the carbohydrates prevent rosette formation by the		
CC	blood cells, this prevents occlusion of capillaries as is seen in		
CC	cerebral malaria caused by Plasmodium falciparum. The products of the		
CC	invention can be used to treat malaria or to vaccinate against it, or		
CC	used to design a model to identify compounds that bind to PfEMP1. The		
CC	carbohydrates, polypeptides and antibodies of the invention can be used		
CC	as a medicament for dissolving the rosettes formed by erythrocytes		
CC	infected by a malaria parasite. The products can also be used for		
CC	detection, diagnosis and drug screening.		
XX			

Seq	Sequence	2228 AA:
Alignment Scores:		
Pred. No.:	9,266-113	Length: 2228
Score:	1723.50	Matches: 770
Percent Similarity:	29.07%	Conservative: 326
Best Local Similarity:	20.42%	Mismatches: 660
Query Match:	8,868	Indels: 2016
DB:	20	Gaps: 134
US-10--087-013-1 (1-10628) x AAW93944 (1-2228)		
QY	AAAAGTCAAGAAATGTTTGGACGTTATGGCAAAATATTAAGACATTCATCAAAATAT	156
Db	13 GUAAPALALYENH1SV1LEUASRGLURHPEG1GLNYSVA1---H1ASRGLV1AH1IS	31
QY	157 -----GCAAAAGACATGTGGATTCGTTGAAAGGGATTTGACGAACGAATTTGCT	210
Db	32 GLYGLV1AL1LYASNP1YVALSERGL1LEUYSGL1SER1LEU1ASER1LEU1	51
QY	211 GGTGGTCCTTACGCGCAAGTAATAGCATATTAATTTATTCATATCCATGTAATTTA	270
Db	52 GLYGL1URH1AL1RHPH1VAL1YS-----	59
QY	271 GATCATATMGAACATATTAATTTACGCTATGATGTG-----AATTCAGA	318
Db	60 -----SerMetGL1URH1GL1SER1YS1UR1TH1GR1U1EU1EG1UAL1AS1SER1YS1ARG	77
QY	319 CATCCTGC-----CATGGTGAACACAAACGATTTGATGAATGAAGAA	366
Db	78 AANPRCVALY1LYASRPG1LYUG1LYASNP1AL1SP1ARG1RHP1ES1VAL1YS1GL1GN	97
QY	367 TCTGATGTGCAAAATGAATCGTATATTAAGAGAAA---AATGATGCTATAGCCTGT	423
Db	98 ALAAG1LYUR1ASPR1LYS-----LYSMEL1YS1CYS1ER1AS1NG1MET1RH1CYS	113
QY	424 GGGCCACCTAGACAGCAATATGTGTGATTAACAACTTGGAGCCTGTAATGATATTAAT	483
Db	114 ALAPRPH1EAL1YD1GL1EN1H1LE1UC1YS1AS1NP1HER1PRO1S1MET1AS1SER1AS1PR	133
QY	484 ACCCAAAATATTT---CATGATTTATGGCAATATACAGTACGCAAAATGCGAAGCT	540
Db	134 SER1E1YS1AL1LYEN1H1AS1R1LEU1EAL1GL1VAL1YS1MET1AL1AL1Y1ST1RG1UG1LY	153
QY	541 GAATCAATTTGTAATATCATCA-----CATAAAGCACTTA---GAC	582
Db	154 G1U1SER11LY1ST1H1N1ST1Y1PR1LY1SER1YS1UR1T1Y1PRO1G1SER1AS1RHP1RO	173
QY	583 GCTTACTGCTGCTGCACGAAGTTTGGCATATAGCTGATATTTGTAAGAGAA1RAGT	642
Db	174 METCYS1TH1MET1LEU1AL1AL1Y1SER1RHP1EAL1AS1PLE1GL1EAL1ATG1GL1Y1A1YD1AS1PR	193
QY	643 ATGCTT-----AAACCAATGTCATCGCAAGA72	672
Db	194 LEUT1LY1EUG1LYASNP1LY1LY1LY1EUG1LY1ASN1GL1Y1SG1UR1H1RG1UAR1G1UL1YS1LEU	213
QY	673 GAAACGGCTCCGACGAGCTTTTAAAGAAATACATGATGCAATGCAAGCT---GAAGTA	729
Db	214 G1UG1N1LY1SER1YS1G1UL11ER1H1E1LY1YS1LE1H1AS1PR1S1NEU1YS1AP1LY1G1UAL1A	233
QY	730 AAAAATGATTAACATCTCGATGATCGAAGTAATATATTAATTAAGAGAGATGCTGG	789
Db	234 G1N1YS1ARG1Y1R1AS1NG1LY1AS1RPG1UAR1PRO1S1RHP1H1E1Y1YS1LEU1ARG1UL1AS1PR1T1PR	253
QY	790 AATGGAATTAAGAAATGAAGTAAGGAGCTATTAACATGTGATGCATCATAT---AAATCT	846
Db	254 THR1AL1ASN1ARG1UL1UR1VAL1TRPG1LY1AL1MET1RH1CYS1ER1YS1G1LEU1EUS1P1ASN1SER	273
QY	847 GGATATTTTATGACATCGAAGAAAGTAATAC-----CCATTAATTTTCAATGCTAA	897
Db	274 SER1YR1RHP1EAL1G1AL1TH1R1CYS1ASN1AS1RHP1E1GL1N1GL1Y1PRO1SER1GL1UR1H1AS1NP1YS	293
QY	898 TGC-----GGCCATTAACAAAGGAAG-----	918

```
Db 294 CysArgCysAspLysAspLysGlyAlaAsnAlaGlyLysProLysAlaGlyAspGlyAsp 313
QY 919 -----GTGCTCAACAATTAGATTATGCTCCCAATATTAGCTGCTGACGAA 969
Db 314 ValThrIleValProThrIlePheAspTyrValProGlnTyrLeuTyrPheGlnGlu 333
QY 970 TGGGAGAAAGAGCTTTGCCGAAAGAAATTTAAATTTGAAAAAGGTCAAGACGCTCTGT 1029
Db 334 TrpAlaGlnAspPheCysAlaGlyLysLysLysLeuGlnAsnLeuGlnLysGlnCys 353
QY 1030 CGT---AATACAAA-----GAACCTTATATTGATGTCATATAGCAGCATGTTGTACG 1080
Db 354 ArgGlyLysAspLysSerAspGlnTyrArgTyrCysSerArgAsnGlyTyrAspCysGlu 373
QY 1081 ACAACTATTTGAAAAAGATTTTGTGATTTGCAATTAATAGCTGCTGCTGCTGACT 1140
Db 374 GlnThrIleSerArgLysGlyLysValArgMetGlyLysGlyCysThrAspCysPhePhe 393
QY 1141 AAATGCAAACTTTTGAAGTTTGATTTGATGGAATCAACAGAACATTTTAAAAA 1200
Db 394 AlacGlySerTyrGlnAsnTrpIleAspAsnGlnArgGlySerGlnPheAspLysGlnLys 413
QY 1201 GAAAAATATGAAAAAGAAATACATATTTTTCAGACATACAAATTTTGCAATAT 1260
Db 414 ---LysTyrThrLysGlnIleSerAspGlyGlyLysArgLysLysArgAlaValGlyGly 432
QY 1261 ATTAAT-----AGTGAATATTTATTAACAATTTTATGAAAAACCTTAAGAAAGCAATAT 1314
Db 433 ThrThrLysTyrGlnGlyLysGlyLysSerPheTyrGlnLysLeuLysAsnAspGlyTyr 452
QY 1315 GCAACTAATGACACTTTTAAATTTACTAATGAAGAAAGATTTGTAA----- 1365
Db 453 GlyThrValAspAlaPheLeuGlyLeuLeuAsnAsnGlnLysAlaCysLysAspIleThr 472
QY 1366 -----GCAGCA 1371
Db 473 AspGlyGlyLysIleAsnPheLysGlnValAsnSerGlyGlyAlaValAlaGlyGlyGly 492
QY 1372 TTACAGAGAGAAAGATATTACTTTTACTACAGTGCCTATGACAAAGGATATTAT 1431
Db 493 SerGlyGlyThrSerGlyAlaSerGlyTyrThrAsp---GlnAsnLysGlyThrPheTyr 511
QY 1432 CGTTCAAGATATTTGCCAAGTGTGCCGACTGGGGGTCAAAATGTGATGTTAAAAATAC 1491
Db 512 ArgSerGlnTyrCysGlnProCysProAspCysGlyValGln----- 525
QY 1492 ACAACAAAACAGATATATGCTGTAACGCTGAAT-----AAT 1530
Db 526 ---HisLysGlyLysAsnGlnTrpGlnArgLysThrLysValLysLysMetArgTyrSer 544
QY 1531 GAAGACTATTAACCTCATGGGTGGAAGCCTACTAATATCACTGCTTATAGTGT 1590
Db 545 LysLeuTyrLysProIleAsnGlyLys-MetValIleLeuLeuLysSerLeuLysValIle 564
QY 1591 AATGAACAAGGTATATTCACAAAAATTA-GAAAAATTTTGT-----AACAGCTC 1640
Db 564 LysAspMetMetIleLeuLysAsnTrpLysGlnPheCysIleThrGlnAsnSerSe 584
QY 1641 A-----ACTAATTACAAAGATRAAAATTAATCAAAAA-- 1671
Db 584 AspGlySerValGlySerValAlaThrThrGlyAlaSerGlyLysAsnSerGlnLysLys 604
QY 1672 -----TGGGATGCTATTTATAGGATGAAAAATTAATTAATGATTAAC 1715
Db 604 SerLeuTyrAspGlnTyrLysCys---TyrLysHisAsnGlnValGlnLysValAsnVal 623
QY 1716 GGAACAAAATACTGAATCAATATGAT----- 1743
Db 623 IGIinglyGlnValGlnGlnAspAspGlnLeuLysGlyAlaGlyLysLeuLysIleLe 643
QY 1744 ---AATCCTAAG----- 1752
Db 643 UPProAsnProLysLysAsnLysGlnValSerGlnAlaLysSerGlnAsnAsnHisAlaAs 663
QY 1753 ----ATAATATCATTTCAATATTTTGAATATAGGGTATACATTTATTAAGGATAC 1808
Db 663 PIIeGlnLysThrPheHisAspPheTyrTyrTrpValAlaHisMetLeuLysAspSe 683
QY 1809 TATTAAAGTGAAATGACAAA---CTTAAACTTGTATTAATTAAT---ACAACACGCAATG 1862
Db 683 rIleHisTrpArgThrLysArgLeuLysSerCysIleSerAspGlyLysThrMetLysCys 703
QY 1863 TATTGATGATGTACAGAAATTCCTTATGTTTGACAGATGGCTTAACAAAAAGAAACA 1922
Db 703 sArgAsnGlyCysAsnLysLysCysAspCysPheGlnLysTrpValLysGlnLysGlnThr 723
QY 1923 AGAATGGAATAGATATAAGAAACTGTCACAAAAAAGAAATATACAGCAATCGATTA 1982
Db 723 rGlnTrpLysProIleLysAspHisPheLysThrGlnGlnGlyIleProGlnLysLysTyr 743
QY 1983 TACTAATATTAATTAATCTTTTGAAGTTATTTTAAAGTTATGATTAACCTGACAA 2042
Db 743 r----- 743
QY 2043 AGATGAAGCAAAATGGAAGAACTTATGAAATATAAAAAGAAAAAATGAGTTTC 2102
Db 744 -----PheThr 745
QY 2103 CAATTGAAAAATATATAGGACTATTTAGAAATGCAATGAACTCTGTTAGACACTT 2162
Db 745 rThrLeuGln-----LeuIleLeuLysLeuGlnPheLe 756
QY 2163 AAAAGAACTGCCACGATATGTAAAGACATAATTAACAAGAAAGATGCAAAACATCCCA 2222
Db 756 uLysGln-----AspThrGlnLysAsnThrGlnLysSerLe 768
QY 2223 TAATGCACACAAACCCTGTGTAACTCTGTGAGGACGCAACCCACTTAATAATAT 2282
Db 768 uAspAla----- 770
QY 2283 AAAAGAAATAGCACATTTCTTTAAAGAGTGCATACGAGAGAAAGCAAGAAATGCTGTCT 2342
Db 771 -----GluGlnAla----- 773
QY 2343 TCATTAATTTGAAAGAAAGACACAGAGGTATATTAACCTGGGGGTAGAGAAAGGA 2402
Db 773 ----- 773
QY 2403 CTTCAGAGCAATTTATGTAGATTAATGATTAACATTTCAATCTTAATCTTGTTTTC 2462
Db 773 ----- 773
QY 2463 AATGAGACATGTGATGGCAAGACAGCAGAGTGATGATACAAACAAGATTTGCTGAGG 2522
Db 773 ----- 773
QY 2523 AACTGAATGGGAAGTGAATCCGAGACACATGCTTAAGATCAGGAAGATTTATATGCC 2582
Db 773 ----- 773
QY 2583 TCCTAGAAAGCAGATATATGTACATCCAAATTTGGAACATTTCAAAACGATATCACCC 2642
Db 774 -----GluGlnLysHisLeuGln----- 780
QY 2643 ACTTAATGTAATATTTGATGATTTAGTTAATTAATTCCTTTTGGGGATTTCTTCT 2702
Db 780 ----- 780
QY 2703 ATCAGCAAAATATGAAGCAAAACAGATTAATACGAATGTATTAAGAAAGATTAACCTTAA 2762
Db 781 -----LysIleLeuLysLeuGlnLysGln---AsnAsnLeuLys 792
QY 2763 GGGCCCCAAAGAAAGTAACGTACCCCAAAACACGACAACTATCTGTGACGTATACGTTA 2822
Db 792 aValValAsnAlaGlyThrGln-----GlnLysThrLeu----- 803
```

QY 2823 CAGTTTTCAGANTATAGGTATATATTTGAGAGAGATCTCTGGGAAAGAAACGGTGA 2882
Db 803 -----
QY 2883 CATGTAAAGCTGCAAGACATTTTGAACCTGTTTGGTAATATACATATGCTACTCAA 2942
Db 804 -MetaspLysLeuLeuasnHis-----
QY 2943 AGCCAAAGAAATGATTAATATATATGATGATGCCCCCAATATTAAATTGAGGAA 3002
Db 810 -----
QY 3003 TTGGTGGAGCATATATAGCAACCAAGATGCGAAGCCAAATGATATATAATATTT 3062
Db 811 -----GluLeuasn-----AspAlaThrLysCysGlyAspCysProLe 823
QY 3063 G-----AAGGATAATTCGGGACACCAATCAACAAAGATGTTATTTGGCATATAGTGA 3116
Db 823 uProGluGluAspLysSerArgLysArgSer-----AlaAs 835
QY 3117 TCATACACCATTTGATGATTATATCCACAAATAATTAAGATGATGACGAATGGGACGA 3176
Db 835 pProSerPro--AspIlePhePro-----
QY 3177 ATGCTACTGCAAGCTGCAAGAAAGAGATATGATTAAGTTGAAGAGAGATCTAGAGATG 3236
Db 844 -----ArgProGluGluLysGluAspAspGlu-----
QY 3237 TAAGATAAGATTAATGCTCAAGGCTGTACGAAGAGAGATGATGATGTTGACGAAGTG 3296
Db 852 -----
QY 3297 CACAGAGCTTGTATGATATATATATATATATAGATTATGAAAGAAATGGAATAT 3356
Db 853 -----AsnGluAspAspAsp-----
QY 3357 AATATCAGATTAATACAAAGAAATTAATGACAGACAAACATCTCTGTTAGTAATAGTGC 3416
Db 858 -----GluAspGluValAlaArgAspAspGluLysThrAlaLysGluThrGluGlySerAl 876
QY 3417 TATTTGAGCTTCAGTACTGCCAAATATCATATAGACAGATGTTATGAAATTTTGTGC 3476
Db 876 aThrAspThrThrThrSerLeu-----
QY 3477 GGAATTTATCCAAACAAATGTGTGCAAAAGTAATAAAGTGTACTAGTGAAGATGC 3536
Db 884 -----AspValCysPr 887
QY 3537 TGTCAATGCTACTTAACACACGATATGAAATGTTGAGCATATCTCCATGATACAGGAA 3596
Db 887 olLeValGlyLysValLeuThr-----
QY 3597 TTTTGTATGATTTGTCACTACCAAAATGAGTTTGTATGTAAGAAAGTGTATGATTA 3656
Db 895 -----LysAspAs 897
QY 3657 CGAATAATATATGCTTTAGAGATTAACACAGACCATGATGCTGCTGTTGTAAG 3716
Db 897 nGluSerLeu-----GlnAsp-----AlaCysSerLeuLysTy 908
QY 3717 TGGATCGAAGCGACAGAGGTACAGATTAACAAAGGAGGAGGAGGAGGAGATAC 3776
Db 908 rGlyLysAsnAsnSerArgLeuGlyTTPArgCysValThrProSerGlyGluProThrTh 928
QY 3777 GGAATGTAAAAACGTGAATGATATATCTTAAGAAAGACATGGAAGAAACAAAGTAGAGA 3836
Db 928 r-----
QY 3837 TTGTATCCAAAAAGATATATGATATCCGATTTGCCAATGGCAATATATTAATTT 3896
Db 929 -----SerSerAspLysAsnGly----- 934

QY 3897 AGTGAAGACCTCGTGTGTATGCCCCCTAGAAAGCAAAAGTTATGCTACTTTCT 3956
Db 935 -----AlaIleCysValProProArgArgArgGluLeuTyLeLysLysI 950
QY 3957 G-----CCAAATGATTAATTAAGAAAGCTTCATCAATC 4022
Db 950 eValAspTrpAlaThrLysThrGluSerProGlnLysSerGluLysAlaPheValGluSe 970
QY 3984 A-----CAATCACAAGTTAATTAAGAAAGCTTCATCAATC 4022
Db 970 rSerGlySerThrThrProProAspSerLysGluLysAlaLeuLysAlaPheValGluSe 990
QY 4023 TGCAGCAGCAGCAACATCTCTCATGATATATTA 4064
Db 990 rAlaIleArgLysThrThrPheLeuThrPheHisArgGlyTyrLysGluLysLysAlaValAl 1010
QY 4065 TAAAGATGTGAAGAAATGAATC----- 4089
Db 1010 aGlnGluLysAlaGlyHisGlyLeuProArgValGluGluLysSerProGluTyArgPr 1030
QY 4090 -GATAAGATTAAGAAAGCAAAATCTCCGCCATTTTGAATCATGTTCTACAC 4148
Db 1030 oGluAspLysLeuLysGluGlyLysIleProAspGlyPheLeuArgGluMetPheTyTh 1050
QY 4149 ATTTGAGATTAATAGATTTTATTTTGAACAGATATATCAAAAGGTCATGCTGAGGG 4208
Db 1050 rLeuGlyAspTyrArgAspIleLeuPhe-----
QY 4209 AAGTAACTTAAAGCAAAATAGATTTCTTTTCAAAAATGCTGACCAAAAATCTCCTAA 4268
Db 1059 -----
QY 4269 TGAAGAAACAGCCAAAGATGTGTGACAGACATAGTCATGATATGGAAGCTATGCT 4328
Db 1059 -----
QY 4329 ATGTGCTACTATAAAATTTGGGGCAAAAAGATGATTTTACGAAACATACGGTTACAA 4388
Db 1059 -----
QY 4389 CAACTCAATTTAGTGCAAAAGACACACCTTTGGAGAAATTTGCCAAGACCCAGTT 4448
Db 1059 -----
QY 4449 TTTAGATGCTTAACCGAATGTACAGACTATTTGCTATACAGCAAAATATTTGAA 4508
Db 1059 -----
QY 4509 GGATGTGCAAGAAATGTAATGTAATGACCAATTAAGTGTGATACAGATGTATTA 4568
Db 1059 -----
QY 4569 GAAATGCGAGCATACGTTAATATATGAAAAAAGAGTGGATTCACAAAGTAA 4628
Db 1059 -----
QY 4629 ATATTAAGAGATGAGCGACCAAAAAGATTCATGATAGACAACACATGTGTAAATGT 4688
Db 1059 -----
QY 4689 TACAGACTATACGTGAACGAATGCAACAGATTTACTTGAACAGAAATTTACTGCTAGTTG 4748
Db 1060 -----SerLysSerAsnAspThr-----ThrSerValSe 1069
QY 4749 TGGTATACGCTGGAAGCTCTGTGGTATACAAAGAAATATATCAATTTGTATAAAAACA 4808
Db 1069 rLysAspThrProSerSerSerAsn----- 1077
QY 4809 GCGTTACTATGATGCCCAACAAATGTGGTGCACAAAATTTATGAAATGACGACAA 4868
Db 1077 ----- 1077
QY 4869 ATATACTAATTTCCAGTAAAGATTAAGTGCAAAGATTAAGTAAAGAGCAACACAGG 4928

Db	1077	-----	-----	-----	1077
QY	4929	TGCTATTAGTGGCAAAACAAAGTCTAATACATACATATCTTGAAAGATGACTGA	4988	----	----
Db	1078	-----	-----	-----	1084
QY	4989	AGATGTGCTTTTTCCTCTCTGCTGCACTACGTATATGTTTCTATGCATTGGATGCCATTA	5048	----	----
Db	1084	-----	-----	-----	1084
QY	5049	TACAGATCCAGAAAGTTAAAGATGAAATATGGGTTGCGAAAAAGATTGATGGAAGTGGCGC	5108	----	----
Db	1085	-----	-----	-----	1090
QY	5109	AACGGAAGGTACATATTTGGGTCAATACATACAAAGAAAAAGAAAAAGAAAAATTA	5168	----	----
Db	1090	-----	-----	-----	1098
QY	5169	AACGTGCGATCCGACACAAATATCTTATAGAGTCCGCGCTTAACTGCTATGAATATAG	5228	----	----
Db	1099	-----	-----	-----	1101
QY	5229	TTTTATGATTTAAGATATATATCTAGGTATGTATATTTGGAAGATGAAAAACAAA	5288	----	----
Db	1101	-----	-----	-----	1101
QY	5289	GACCGAGAAAAATTTGAAGAAATATTTAACAAAAATGCAACATCAGTGGCAAGAG	5348	----	----
Db	1102	-----	-----	-----	1116
QY	5349	TGATAGTACTACAGAAATCCCGGTAGTACTGCGGCAAAATTTTCTGGAACGAAAAATA	5408	----	----
Db	1116	-----	-----	-----	1131
QY	5409	GGAATGTGTGAGAACGACATGATATCGGGTACAAACGTGATGCGATGATGAATAG	5468	----	----
Db	1131	-----	-----	-----	1142
QY	5469	TGGAATATGTGCAAGAAAGTATGAATCTMAAAAAATGTGTTGTATCTTCAGATGA	5528	----	----
Db	1142	-----	-----	-----	1142
QY	5529	TGATATCTATGAGGAAAAATCGCATGAGGTACTGCTATCATGTTCTTCATGTT	5588	----	----
Db	1142	-----	-----	-----	1142
QY	5589	TGCCGAATGGGTTGAAGATTTTTCACAAACATAAAGAAAAAGAAATTTGTTAG	5648	----	----
Db	1142	-----	-----	-----	1142
QY	5649	GCGGTGAATGATTAATCTTGTGGTATATGAGATTAAGAAAGAAATGTACAGATGC	5708	----	----
Db	1142	-----	-----	-----	1142
QY	5709	GTTGACACATATATAAAATTTATTTAGTGTGAGGAACACAGTATGAAAAACAAATCA	5768	----	----
Db	1142	-----	-----	-----	1142
QY	5769	AAATATGTTGAGATTAAGACAAATATATATTCGAGATCCTGTCGCAAAAGATGAGA	5828	----	----
Db	1143	-----	-----	-----	1153
QY	5829	GGAAGCTCGGAATATTTAGACAAACAAATTTAAAAAAATTTGTGAAAAATTAAGTGGAGA	5888	----	----
Db	1153	-----	-----	-----	1153
QY	5889	TTTGAAATTAATAGTATGAAGATGTGTCCACACAGCATTAACGATGATAGTATGCA	5948	----	----
Db	1153	-----	-----	-----	1153
QY	5949	AAATATGCCCCGATCATTTAGACGATGAACCAAAAGAGTTGAAGAAAGTTAATTGTCA	6008	----	----

Db	1153	----	1153	-----	1153
OY	6009	AGTGCCACGAGGTCCACCACGCTAGCAGMGGAAACACCCTACACACGGGTATCACTGAT			6068
Db	1153	----	1153	-----	1153
OY	6069	ATCAAAAAGCGACGCTCGAATAAAAAAGAGCGAAACAGCGCGCGCTACAAAACAGCGCA			6128
Db	1154	----	1154	-----	1157
OY	6129	AAAAAGTGAATAATCTAAACAAGAAATGGGAGCACAAACGACCAGACGACACA			6188
Db	1157	nlgyi1lelunsn			1161
OY	6189	ACAAAACGAGAAAACGAACTCAACACGACACACACAGAAATCTGACGTGGGCACAATGCT			6248
Db	1161	----	1161	-----	1161
OY	6249	AAAGGCCATTCCTTCGAAATTAACCAAGATAGCAGGGGTGAATAGAGGTTGTAATCCAA			6308
Db	1162	-----	1162	-----	1174
OY	6309	AACGTATGACACATATCTTAATAGGGGTGTATTTAGTAAAGCTAAAGAAATGAAAA			6368
Db	1174	S-----	1174	-----	1174
OY	6369	TGGCATATGTATGCCCTGCTAGGAGAAAAAATTTATGTAATTAATATACATATTAA			6428
Db	1175	-----	1175	-----	1182
OY	6429	TTATGAAACCTGMAAATAAGCGTGCACATGATATAAAGAGCGCTTTATTAATGTGCAGC			6488
Db	1182	nvallysleuasp			1186
OY	6489	AATGAAACCTCAATTTTGTGTGTTAAATATATATTGAAAACTCTGCAGCAGAAAAA			6548
Db	1187	-----	1187	-----	1188
OY	6549	AATGCAAAATGGAACAATTCAGATGATTTAAAGAAATGATATATACATATGCGA			6608
Db	1189	-----	1189	-----	1197
OY	6609	TTATTAAGATATGTCTTTTGGAACTGATATTCTTAATGTAATAAAAAATTAATACTGTAC			6668
Db	1198	-----	1198	-----	1203
OY	6669	AAATAGTGTACACCACTTCATCATGAAATATATAGAAAAACGAGATAAAAAAAAGA			6728
Db	1204	AsnThrProthr			1208
OY	6729	TGAAGAAATACGTAATAATTTTGGAGAAAAATMAAAATTTATTTGGAGAGAAATGAT			6788
Db	1208	-----	1208	-----	1208
OY	6789	ATATGATTAATCTATATCTCCACAGACGAAAAAGAAAAATTAGAGATTA			6848
Db	1208	-----	1208	-----	1208
OY	6849	CCAGTACATATGACATGACAAACTGACGCCCTTCCTTGAGAGATTGTGTAAGAGGCCCA			6908
Db	1209	-----	1209	-----	1217
OY	6909	ATTTTTGAGATGTTCAACAGAAATGGCGAGAGAAATTTGTAATTAAGAGAGAACAGTT			6968
Db	1217	RTYpHeRtgrtPheglunlurtpelgyluserPhecysatgguatgylsYsArgle			1237
OY	6969	GTTAAAAATTTGGAGCGCGCTGTATGAGAAATATGATGTATGTGT			7013
Db	1237	uLysclnlelysYsAlpsYsYsValglu			1255
OY	7014	TAAATGCGGTAAACACACAGAA			7035
Db	1255	IGlYAspIygluulAcYsAspserlleserThrHisAspTYserThrValProserPh			1275


```

QY 7036 -----TGTCAAGGCGTGTGTAACATATCAAAATTTATTAAGAGTGAA 7082
Db 1275 easnCySProglYcysglYshsCysSerSerYArGlyTrpLleGlulnrgYsly 1295
QY 7083 AACGATATGAAGACAAAGAGAAAGTCAAAAGAGTAAAGTGGCAAAAGTATA 7142
Db 1295 sllleGlupheHlelYsglnSerlnalArYglYln----- 1307
QY 7143 GGATATCTTCCTGAGAAAGACATAGAGAGCAACATGTGTCATGATATTATAA 7202
Db 1308 -----GlnYsThrAspAlaThrArgsnasnGlyAs 1318
QY 7203 CATGAATTAAGAATATATGTGCAATTAAGATTGTTCTTGATGCAAAAACCTTC 7262
Db 1318 nThrPheAspLysGlupheCys----- 1325
QY 7263 ACAATACCAAAAAACAACACATCATCATCCGATGCTAATGATATGCCAATTC 7322
Db 1326 -----LysThr 1327
QY 7323 GCTGATATATCTTCGAGAAATTTAACAAGTGTAGTCTGAACTTTCAAAAAGG 7382
Db 1327 rleuGlulnThrTrpProAsp----- 1333
QY 7383 ATCTATGATTCATACAAAAAATTTACTGAACTAAATACCTAATGATTTGTAGAGA 7442
Db 1333 ----- 1333
QY 7443 AGCAGATATTTATTTCTTAAGAGACAGAAATATATGATATTACCTGAGAGAAA 7502
Db 1334 -AlaAlaLysPheLeuGlulnArg----- 1340
QY 7503 ATTTATACCTATTTGACTCTACAAAGAGAAAGAACTAAATAGTTGACTAATATAA 7562
Db 1341 -----LeuLysasnGl 1344
QY 7563 TCCTTCGATCTCTAAGAAACCTTATACCTGATATAATATAGAGAGAAACCCCTG 7622
Db 1344 yProCysLysThrAsnLysGlulnYrGlyAsp----- 1355
QY 7623 TCAAAATAGAGAAAGAAATCGTTTAAGGTAGATTAATGAGAAATTTACAAAATTC 7682
Db 1356 -----AspLleAspPheGlu-----LysAspSe 1363
QY 7683 AAAGTTCTATGAGAGAAAGAAAGATATGTACTCT-----CCAAGAGAGAAATATGTG 7739
Db 1363 rlyshThrPheGlnHlsThrGlu--TyrCysglYrProCysProlYsPheLysThrAsnCy 1382
QY 7740 CTTAAGATTTAGATGAATTAATTAAGAACTTAAGACTTAAGTACTAATTTACTCTTAA 7799
Db 1382 s----- 1382
QY 7800 AATGTTGTCGACTGACACCAATGAAGATGACATATATAAAACCTCACTCAGA 7859
Db 1383 -----G1 1383
QY 7860 GAACGGG-----TGCCATGATCCATATATGTAT----- 7890
Db 1383 nasnLysnCySglYValSerGlyLeuAsnGlyAsnCySaspgLysAspLysSerLleAs 1403
QY 7861 -----ACTATGAATATATAGTTTCCTGATGCTGGGTGACATAGTAGAGG 7934
Db 1403 PALALysGlulLeAlaLysMetArGSerSerThrThrAspAla-----ValMetArgVa 1421
QY 7935 AACGATATGTAGCAATTTGGTGGTACTTACCTCCCTAGAAATTAATATATTAAGT 7994
Db 1421 lSerAsp----- 1423
QY 7995 TTTGAATACATATATGAGAAATGAGAAATATAAAGTAGAATATAATACACGA 8054
Db 1424 -----AsnAs 1425

QY 8055 TGTACAAAGCTT-----CGTTCGTGCTGGTGGATGCTAAT----- 8091
Db 1425 pThrAsnThrPheGlnGlyAspAspLysAspAlaCysGlnHlsAlaAsnLlePheHy 1445
QY 8092 -----AAGAAAGATATTTGGAAGCAATGACGTGCAAGCACCAGCAAGATGCAAAACT 8144
Db 1445 sglYlLeAlrGlyAspValTrpLysCys----- 1454
QY 8145 TTTTGAAGAAAGAGAGATGATGATTTGAA-----CGCATACATTAATACAAATATA 8198
Db 1455 -----GlyTyrValCysglYValAspLleCysglulnThrAsnLleAsnGlnAr 1471
QY 8199 GGTGGACATTAAGACGATCCACCTGTTGATGATATATATACCTCAACGGTTT----- 8250
Db 1471 gThrAspGlyLys-----GlnYrLleGlnLleAlaGlnAlaLeuPheHy 1485
QY 8251 -CGATGGATGACTGAATGCTCTGAATATATTAATTAAGACACTGATGGAAGATTTGAAA 8309
Db 1485 sarGtrPvalGluAsnPhe-----LeuGluAspTyrAsnLy 1497
QY 8310 ATTTAAAAATCATGTGATCCTGT-----AAAACATCTGACAGATGCAAGATGA 8360
Db 1497 sllleAsnAspLyslSerHlsCyslLeLysLysglYglulnYserLysCyslLeAsnG1 1517
QY 8361 TTATGATGAATAATAGTGTGACACTGTAAACGATGTCAAGAAATTAATAATTTTGT 8420
Db 1517 YCysglulnYasn-----SerLysCysLeuGlulnYsrTrpLleGlulnYsly----- 1533
QY 8421 TCTTAATGGAATTCCTATTCGATATACATCAATTAATTAATACAAAGATTTGTAA-- 8478
Db 1533 eAlaGlulnTrpGluAsnLle-----LysLysrThrPheAsnAspGlnYrGluAs 1549
QY 8479 -----CAACCAATATATATACAAAATCTCTACTATGATCATGATCAATTTGTACA 8531
Db 1549 nLysAspGlnProAspYrAsnValYsSerLle-----LeuGlulnLeuLlePr 1566
QY 8532 AAAGTGAAGAACTTTAAAGATGAATGTTCTGTGAGAGCTTTTCTGAATATCTTCATGA 8591
Db 1566 olYsllleAlaValAlaAsnAspGlnAspAsnValLllyLysLeuCyValPhe-----G1 1584
QY 8592 AACAGTATGTTTGAATTTAATTAATTAATGAAGAAATGATGCTTCCATATACCAAC 8651
Db 1584 uAsnSerLysglYsThrLeuLleSerAsnThrGlnAsnAsnLysglulnAsn----- 1601
QY 8652 ATATGCTTTCGAGAGAAACACCAAAAGTATTAAGAACTTCGACTGACACTCTTC 8711
Db 1602 -----AspAlaLleAspCysSerLeuLysly 1610
QY 8712 TAAAGAT--CATTTGATATATGT-----CTTACCGATCAAAACAAAGATGATG 8759
Db 1610 sleuGlYValLysAlaLysAsnCySProglYlYsProSerGlyglulnYsglnSerAspCy 1630
QY 8760 TAAAGAAATTAACAACCTTTTACCTCTGCTCGAAGAAATGATTAATATATCTTGATAA 8819
Db 1630 slysglulnProProleu-----ProAspGlnGluAspGlnAsn----- 1643
QY 8820 TTGAGACGATACCTGTCTTAAATAGTCAGATGAATTAAGAGAGTGTATGCTCTCC 8879
Db 1644 -----ProglulnAsnThrLeuGlulnProPr 1652
QY 8880 AAGAAAGAACATTTATGACAGACCTATACCTGATATATATTAAGAAAGGTGATAA 8939
Db 1652 olYs-----PheYsPro--ProThrThrGlnProProglulnYsly----- 1666
QY 8940 AGAAATTTAAAAAAAACCTTCTACTCTGCTTCAGTCAAGCAACATTTGATGTC 8999
Db 1667 -----GlyGlnThrCysglYAs 1672
QY 9000 AAAATATAATTCGAAGACAGTTGTGCTTTGAGCAATGAATATATATGATACAGATTA 9059
Db 1672 nLysglulnLysLysAspLuln----- 1679
QY 9060 TTCGATATATTAAGAAACATATGATGACACTTCATTTCTGAAAAATTAATAA 9119

```

[illegible]

PI BaruchDI, Howard RJ, Pasloske BL;
XX
XX WPI: 1996-467376/49.
DR N-PSDB; AAT41853.
XX
XX
PT New Plasmodium falciparum erythrocyte membrane proteins - used to
PT develop products for the diagnosis, treatment or prevention of
PT malaria parasite infections
XX
XX
PS Claim 1, Figure 12; 149pp; English.
XX

A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte membrane protein 1 (PfEMP1) or active fragments or analogues of that protein can be used in the treatment or prevention of symptoms of that malaria parasite infection. The polypeptides can inhibit, block or reverse the sequestration of erythrocytes in patients suffering from malaria. Nucleic acids derived from the PfEMP1 gene can be used as probes and primers to identify a Plasmodium falciparum parasite, the CC primers used to generate characteristic amplification patterns from different P. falciparum strains. Antibodies specifically immunoreactive with the PfEMP1 polypeptide or its fragments may be used in diagnosis of malaria infection. This is a truncated PfEMP1 protein of the MC type of Plasmodium falciparum. The full length PfEMP1 protein is given in AA000384.

SQ Sequence 1726 AA;

Alignment Scores:

Pred. No.	4.91e-79	Length:	1726
Score:	1244.50	Matches:	501
Percent Similarity:	34.50%	Conservative:	276
Best Local Similarity:	22.2%	Mismatches:	632
Query Match:	6.40%	Indels:	843
DB:	17	Gaps:	88

05-10-08/-013-1 (1-10628) X AA#00385 (1-1726)

QY 97 AAAAGCGCAAGAAAGTTTGGAGCGTTATGCCAAAAATATA----- 135
 ||| ||| : : : : : ||| : : : : :
 Db 13 LysAspAlaIleuHsAlaIleuAspArgIleGlyGluIleValTyrLysGluLysValGlu 32
 QY 139 AGCATCCATCAAAATATGCAGAAAGAACATGTGGATTCGTGAAGGGCATTTGACGAA 198
 ||||| ||| ||| ||| : : : ||| : : : ||| : : :
 Db 33 AsnAspAlaGluLysTyrLysLys-----AlaIleuLysGlyAsnIleuGlu 48
 QY 199 GCAGATTTTCGTGGTGCCTCTTACGCCAGTAATAGCATTAATTATTATTCACAT 258
 ||||| ||| : : : : : ||| : : : ||| |||
 Db 49 AlaLysGlyIleGlyGluLeuAlaSer-----Proasn 60
 QY 259 CCATGTAATTTATGATCATAGCAACACTAATTAATTCGGTATGATGTAATTTGAGA 318
 ||||| ||| : : : : : ||| ||| |||
 Db 61 ProCysLysLeu-----TyrGluAspTyrTyrAsnAsnArgLeuLys-----ArgLysArg 77
 QY 319 CATCTTCGCCATGTGAGAGCAAAACCCGATTTGATGAAGATGAACATCGATGTGCGA 378
 : : : : : ||| : : : : : ||| : : : : : ||| : : : |||
 Db 78 TyrProCysAlaAsnArgGlnThrValArgPheSerAspGluTyrGlyGlyCysThr 97
 QY 379 ---AATAAATACGTAATTTAAAGAGAAAATGATGCTATA---GCCTGTGCCGCACT 432
 ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : |||
 Db 98 PheAsnArgIleLysAspSerGluAsnAspAsnSerLeuGluYalaIleProTyr 117
 QY 433 AGAAGACACATATGTGTGATTAAGAACTGGAGACGCTTAATGATATAATACCCAAAT 492
 ||||| ||| : : : ||||| ||||| ||| : : : ||| : : : |||
 Db 118 ArgArgGluLeuHisLeuCysAspTyrAsnIleuGluLysMetGluLysThrSerThrThrLys 137
 QY 493 AATTCGATTTATTGGAAAATGTAAGTAGTACAGCAAAATATAGCAAGGTGATCAATGTT 552
 ---||| ||||| : : : ||| : : : ||||| ||||| ||| : : : |||
 Db 138 ---HisGlyLeuLeuLeuAspValCysMetAlaIleLysTyrGluGlyAspSerIleLys 156
 QY 553 AATATATCT---CCACATTAAGAACT-----TCAGACGCTTGACAGCT 594
 : : : : : ||||| : : : ||| ||| |||||
 Db 157 ThrHisTyrThrLysHisGluLeuThrAsnProAspThrLysSerGluLeuCysThrIle 176

OY	595	CTTGCAACAAGTTTGGCAGATATGAGCGATTGTGTAAAGGAATAGTATGCTT-----	648
Dd	177	LeuAlaIagSerPheAlaAspIleGlyAspIleValaTrgIyLysAspLeuTrIeugly	196
OY	649	-----AACCAAATGTCCATGCACAAAGTAGAAMAACGGGCTCCGAGAGGTTTC	696
Dd	197	TyrAspAspLysGIuLysAspGLuaRgLysLysLeuGlnAsnLeuIleThIlePhe	216
OY	697	AAGAATAATACATGTATGGATATGGAAGATGAAGTAATAAAAATGATTAACAATCTGATGATCT	756
Dd	217	LysLysIleHISgluAsnLeuGlyThGlaspAlaLysAspHisTryLysLysAspGlu	236
OY	757	GCAATTTATTTAATTAAGGAAGACATGGTCGATGTGATTAAGTAATTAAGTATGGAA	816
Dd	237	GluAsnTyrrYGlnLeuarGluAspTrpTrPhalaAsnArgSerThrValTrpLys	256
OY	817	GCTTAACAATGTATGCATCATATAAATTCGATATTTTATGCATACAGAAAGTAATACA	876
Dd	257	AlaIleMrCySHIsalAgLyGluSerAspLysTrpPheArgLysThrCysSercLy	276
OY	877	CCATTAATTTTCAATTCCT--AAATGGCGGCATAAACACAGAAAG--	918
Dd	277	GluTrpThrAspAspLysCysArgCysLysAspGlnGluLyLysAsnGluThrAsnGlu	296
OY	919	GTTCTCTACCAATTAAGATTATGTCCCTCAATATTTACGTTGCTTCGACGAATGGGAGAA	978
Dd	297	ValProThrTyrrPheAspTyrrValProGlnTyrrLeuarGrpHeGlnGluTrpAlaGlu	316
OY	979	GAGTTTGGCGAAAAAGAAATATTAAATGAAAAAGGTCAAGACCTCGTGTATATAC	1038
Dd	317	AspPheCysArgLysAspTrgLysLysIleGluAsnAlaIleLysAsnCySargLyGlu	336
OY	1039	AAA-----GAACGCTTAATTTGTACTCATATGACATATGACATGATGTGCACACPTATGG	1092
Dd	337	LysLysLysAsnGluArg--TyrrCysAspLeuAsnGlyTyrrAsnCysGlnGluThrAlaArg	355
OY	1093	AAAAAAGATTTTGGATTGGATTAATTAAGTACTGACTGTTCGACTAATGCAAACTT	1155
Dd	356	GlyValaGluIlePheValLysGlyAspAspCysHisLysCysSerValAlaLacysAspArg	375
OY	1155	TTTTGAATTTGGTTTGGAGATCAACAAGAAGCATTTTAAAAAANAAGAAANATATGA	1212
Dd	376	PheValLysTrpIleAspAsnGlnArgLysGluPheAspLysGlnLysLysTyrrAsp	395
OY	1213	AAAGAAATACAA-----TCATTTATATGACGATACAAATTTGCTCAATAT	1266
Dd	396	GluGluIleAsnLysThrHisGlyTrpThrIleThrThlGlyAsnGlyLysIleAsn	415
OY	1261	ATTATATAGTAAATATTATAACAATTTTATGAAAACTTATAGCAACCAATATGACACT	1320
Dd	416	Leu-----TyrrAlaGlyHisPheTyrrLysIleLeuLysLys--TyrrTyrrProThr	431
OY	1321	AATGACACTTTTAAATTTACTTAATTAAGAAGAAAGTATGTMA-----GGAGATTA	1374
Dd	432	ValAspLysSerLeuGlnLysLeuAsnAspGluAlaIleCysLysLysProboAsnAl	451
OY	1375	CCAGGAGAAAGATTTACTTTTACTAACACTGCTGATGACAAAGGATATTTATTCGT	1434
Dd	452	GlyAsnGlnLysAlaSerThrValAspPheAsnAnGlnValAsnThrThPheSerHis	471
OY	1435	TCAGAAATATTGCCAATGTGTCGCCGACTCGGGGTCAATATGATGAT-----GGT	1482
Dd	472	ThrThrTyrrCysGlnLacysProTrpCysGlyAlaGlnLysGluLysAsnGlyGly	491
OY	1483	ATTAATATACACACAATCA---GATTAATGATCCTGAACGCTGAATATTAAGACTAT	1539
Dd	492	TrpLysAlaLysGluLysSerCysAlaLysLysGluArgIlePheAsnLysGlu--	510
OY	1540	AAACCTCATGGGGGTGAAGCCTCATATACATGTCCTTTATAGCGTAAATGAACAA	1599
Dd	511	-----AsnSerThrAspIleLysIleLeuThrProGluLysGlyArg	524
OY	1600	GGTGATATTACACAAAATTTTAAATAATTTTGTAAACAGCTCACTAATTTACAAAGATAAA	1659

```
Db 525 SerLysThrLeuGlnLysLeuLysThrPheCysLysAspGlyGlnLysIleLys----- 542
QY 1660 AATATCAAAAAATGGATGCTATTTAAGATGAAATATAATATGATGTAACAGTGA 1719
Db 543 ---AspAspIleThrPheCysHisIleAspAspAsnGly-----ThrAsp 556
QY 1720 CAAATATCTGAATCAATATATGAT-----AAT 1746
Db 557 AspGlnThrAspAspSerAspAsnAspCysValLeuGlyAspIlePheLysLeuThrLysGln 576
QY 1747 CCTAATATATCATTTCTAATATTTTGAATATATGGTTACATATTTATTAAGCAT 1806
Db 577 AspLysIleMetSerLysAsnAlaPhePheIlePheIlePheIlePheIlePheIle 596
QY 1807 ACTATTAAGTGAATGAACTTAAACCTGATTAAT---AATCAACACGACATTTGT 1863
Db 597 SerIleLysIlePhePhePhePhePhePhePhePhePhePhePhePhePhePhePhe 616
QY 1864 ATTGATGAATGTATACAAATGCTTATGTTTACAGATGCGTTAAAGCAAAAGAGA 1923
Db 617 IleLysGlyCysAsnLysCysIleCysPheGlnLysIlePheGlnLysIleLysThr 636
QY 1924 GAATGAATAGTATTAAGAACTGTCACAAAAAAGATATACAGCATGATTTAT 1983
Db 637 GlutPheLysIleLysAspHisPheArgLysGlnLysAspIlePheLysAspIle 655
QY 1984 AGTAATATTAATATCTTTTGAAGTTATTTTAAAGTATGATTAACCTTGACAA 2043
Db 655 ----- 655
QY 2044 GATGAAGCAAAATGAAAGAACTTATGAAATATTAAGAAAAAAGAAATGAGTTTGC 2103
Db 655 ----- 655
QY 2104 AATTGGAATTAATAGCGATTTAGAGATGCAATA---GACTCTTGTAGAT 2157
Db 656 -----ThrHisAspAspPheLeuGlnThrLeuMetLysAspLeuLeuGln 672
QY 2158 CACTTAAAGAACTGCCAGATATGTAAGACATATACAAAGCAAGCATGTGAAGA 2217
Db 673 IleIleGlnAspThrTyr-----GlyAspAlaAsnGln----- 683
QY 2218 TCCCATATATGCAACAAACCCGCTGTAAACCTGTGAGGACGCAACCCACTAAA 2277
Db 683 ----- 683
QY 2278 AATATTAAGAAATATACACATCTTTAAAGAGATGATACAGACAGCAAAATCGT 2337
Db 684 ---IleLysArgIle-----GluAlaLeuLeuGlnIleAla----- 694
QY 2338 GGTCTTCATTAATTAAGAAAGCAACAGATATATATTAAGCTGGGGGTAGAGAG 2397
Db 695 GlyValGlyCysIleAspPheAlaAlaLeuAlaGlyLeuThrLysGlyPheValAla 714
QY 2398 AAGCACTTCAAGACAAATTAATGATATTAAGATTAACATTTCTAATCGTAATCTTGT 2457
Db 715 Gln-----LysAspThrThrIleAspLysLeuLeuGlnHisGlnLys----- 729
QY 2458 TTTTCAATGACCATGTATGTCGAAGGACAGGTGATGCTATACAAACAGATTTGTC 2517
Db 729 ----- 729
QY 2518 GTAGAACTGAATGGAGAGTGCATCCGAGACACATGCTTAAGATCAGAAAGATTAAT 2577
Db 730 -----GluAlaAspLysCysLeuLysThrHisThrAspAspThr 742
QY 2578 ATGCCTCTCAGAAAGACACATATATGTACATCCAAATTTGGAACATTTACAAACGATAT 2637
Db 743 CysProProGln-----GluAsp 748
QY 2638 CACCACTTAATAGTAAATATTTGATGATTTAGTTAATATTCCTTTTGGGGGATGTT 2697
Db ----- 2697

Db 749 ArgSerVal----- 751
QY 2698 CTTCTATCAGCAAAATATGAGCAACAGATATATACGATGATATTAAGAAAAAGATATAC 2757
Db 752 -----AlaArgSerGlnSerAlaThrVal----- 759
QY 2758 CTAAAGGCCCCAGAACATTAACCTGACCCCAAAACACAGACAACTATCTGACCTATA 2817
Db 760 -----ProSerProAlaAspPheLys----- 767
QY 2818 CGTTACAGTTTTCAGATATAGTGATATTAATTCGAGGAAGAGATCTGGGAAAGAAC 2877
Db 767 ----- 767
QY 2878 GGTGATGATGTAAGCTGCAAGGACATTTGAAACTGTTTTGGTATATATACATAAGTCA 2937
Db 767 ----- 767
QY 2938 CTCGAAGCAAGAAAGAAATGATATTAATTAATGATGATCCCCCAAAATTTAAATGAGC 2997
Db 767 ----- 767
QY 2998 GAAATTTGGTGGAACTATAGAGCCAAAGTATGGGAAGCCATGAATGTATATATAA 3057
Db 767 ----- 767
QY 3058 TATTTGAAGATTAATCGGGACCAATCAACAAAGTATATTCGGGATATAGTAT 3117
Db 767 ----- 767
QY 3118 CATACACATTTGATGATTAATATCCCAAAATTAATGATGATGATGATGATGATGAT 3177
Db 767 ----- 767
QY 3178 TGCTACTGCAAGTGCAGAAAAAGAGTATGATTAAGTGAAGAGAAAGTGAAGAGTGT 3237
Db 767 ----- 767
QY 3238 AAGGATTAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3297
Db 767 ----- 767
QY 3298 ACAGAAAGCTTTGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 3357
Db 767 ----- 767
QY 3358 ATATCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3417
Db 767 ----- 767
QY 3418 ATTGAAGCTTCCAGTACTGCCAAAAATCATATATGACAGAAATGTTATTTTGTGCG 3477
Db 768 -----AlaThrGlnGlnValAlaAspAlaAsn----- 775
QY 3478 GAATTAATCAACAAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 3537
Db 776 -----AlaSerSerAspAspGln----- 781
QY 3538 GTCATTTGATTAACACCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 3597
Db 781 ----- 781
QY 3598 TTTGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3657
Db 782 ---AspAspPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 798
QY 3658 GAAATTAATGCTTATAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3717
Db 799 AlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 816
QY 3718 GATTCGAACCCGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3777
Db 817 ProSerProProLysThrThrGlnAspGlyValLysProAlaSerGlnGlnLysAspVal 836
```


Db 563 AlaAlaThrAlaGlyGlyGlyAlaAlaGlyAlaAlaAlaGlyThr----- 577
QY 2542 CCGGAAACATGCGTAAAGATCAAGAGATGTTATGCTCCTAGAAAGACGACATATA 2601
Db 577 ----- 577
QY 2602 TGTACATCCATTGGAACATTTACAAAGCGATGATCCCACTTAATGTTATTGTT 2661
Db 577 ----- 577
QY 2662 GATGATTAGTTAATAATTCCTTTTGGGGATGTTCTTATCAGCAAAATATGAAGA 2721
Db 577 ----- 577
QY 2722 AACAGATATACGAATGTATAAAGAAAGAAATACCTAAAGGCCCAAGAAATACCT 2781
Db 577 ----- 577
QY 2782 GACCCAAACACACGACACTATCTGCGAGCTATACGTTTACAGTTTGCAGATATAGT 2841
578 -----ThrrhrAlaCysGlyAlaAlaAlaAlaAlaCys----- 589
QY 2842 GATATATGTGAGGAAGATCTCTGGAAAGAAAGGTGACATGTTAAAGTGCAGAGA 2901
Db 589 ----- 589
QY 2902 CATTTGGAACCTGTTTTGGTAATATACATACCTCAAGGCAAGAAATGATATA 2961
Db 589 ----- 589
QY 2962 TATATGATGATGCCCCCAATATTTAAATTGAGGAAATTTGGTGGAACTAATAGA 3021
590 -----AlaGlyAla 592
QY 3022 GCCAAAGTATGGAGCCATGAATGTGATATAAATTTGAAGATATAATGGGACAC 3081
Db 593 AlaAlaAlaGlyAlaThrThrGlyCys----- 601
QY 3082 CAATCAACACAGACTTATGCGGATATAGATCATACCACTTGATGATATATATC 3141
602 -----AlaAlaAlaAlaCysGly----- 607
QY 3142 CCACAAAATTAAGATGATGACCGAATGGCAGATGTACTGCAAGTGCAGAAAAAG 3201
607 ----- 607
QY 3202 GAGTATGATATAAGTGAAGAAAGTGTAAAGATTAAGATATATGTCAAAGGC 3261
607 ----- 607
QY 3262 TGTACAAAGAGAGTGTACAGTGTGTACGAAGTGCACAGAGCTTGTATGAATATAT 3321
608 -----AlaThrAlaThrAlaCysThrAlaCysThrGlyAla----- 619
QY 3322 GATATATAGATTTAGAAAGAAACAATGAATATATATCAGATTAATCAAGAAATTA 3381
619 ----- 619
QY 3382 CATGAACACACAAATGTCTGTAGTATAGTGTATGAAGCTTCAGTACTGCCAA 3441
620 -----AlaThrThrGlyCysAlaThrGlyGlyThrGlyAlaAlaThrThr----- 634
QY 3442 AATCATATAGACAGAAATGTATGAATTTTTCGGAATTAATACCAAAATATGTCGC 3501
634 ----- 634
QY 3502 AAAAGTATATAAGTGTACTAGTATGATGAAGTCTGTCTATGTTGTTACTAACACAGCTAT 3561
635 -----GlyThr----- 636
QY 3562 GAAATGTGTGAGCATATCTCATGATACAGAAATTTTGATGATGTCTGACGTCAGAAAT 3621
636 ----- 636

Db 637 -----CysAlaAla----- 639
QY 3622 GAGTTTGTGATGAAAAAGTATGTTAAGATTAAGAAAAATATGCTTTAGAGATATA 3681
Db 639 ----- 639
QY 3682 CCACAGACCATGATGTCGCTGTGTTGTTAAAGTGCATCGAAACCGACAGGTTACG 3741
640 -----AlaAlaGlyCysAlaAlaCysAlaThrAlaThrCysCysAla 653
QY 3742 ATAAAAAGCAAAAAAGCGAAGAAAGATACGAAATCGAATCAACAGTGAATATA 3801
654 AlaAlaThrGlyAlaThrAla-----AlaAlaCysAlaThr----- 665
QY 3802 CTTAAAGAAAAAGATGGAAGAAAGAAACAGTAGAAGATTGTCATCCAAAAAGATAGTAA 3861
665 ----- 665
QY 3862 GATATCCGATTGGCAATCGCAATATATAATTTAGTGAAGACCCCTGCTGTATAG 3921
666 -----CysCysThr----- 668
QY 3922 CCCCCTAGAAAGCAAAAGTTATGCTATACATTTCTGGCAATGATATAATGAAAAA 3981
669 GlyAlaAlaGlyAlaThrCys----- 676
QY 3982 TTACAAATCACAGTATATTTAAAGAGCTTTCATCAAAATCTCAGCAGCAGAAACATTC 4041
677 -----AlaAlaCysAlaGlyAlaAlaAlaAlaThr----- 686
QY 4042 TTCTCATGTTATTTATTAAGTAAGATGTGAAGAAATGAACTCGATTAAGATTA 4101
686 ----- 686
QY 4102 AAAGAAAGCAAAATTCCTCCGATTTTGAATCCATGTTCTACACATTTGGAGATTA 4161
686 ----- 686
QY 4162 AGAGATTTTATTTTGGACAGATATATCAAAAGTCAATGAGGAGAACTAACTAAAA 4221
687 -----AlaAlaThrCysCysAlaThrGlyCysAlaAlaAla----- 628
QY 4222 GACCAATATGATTCCTTTCAAAAAATGTGACCAAAAAATCTCTAATGGAAGAACGCG 4281
698 ----- 698
QY 4282 CAAGAAATGTGACAGAACATAGTATGAGATATGGAAGCTATGCTATGTCAGTATGA 4341
699 -----CysThrThrCysAlaAlaThr----- 705
QY 4342 AAAATTTGGGCAAAAAAGATGATTTTACCGAAAGTACGCTTACAAACAGTCAAAATTT 4401
706 AlaThrGlyAla-----Thr-ThrAlaThrAlaAlaThrAlaCysThrAla 720
QY 4402 AGTGACAAAAGCACCACTTTGAGAGATTTGCCAAAGCAGCCAGTTTTCAGATGCTTA 4461
720 AlaThr----- 722
QY 4462 ACCGAATGTGACGACGACATATGCTATATACAGACAAAAAATATTTGAAGATGTGAGAA 4521
723 -----GlyThrThrAla-Cys----- 727
QY 4522 AAATGTAAGTCAAAATGCAATTTGAAGTGTGATACAGAAATGTAATAAGAAATCGAGAG 4581
728 ThrCysAlaThrGlyGlyThrThrThrGlyGlyThrCys----- 741
QY 4582 TACGTTAATATATGAAAAAAGAGTGCATTCACAAAGTAAATATTACAGAGAT 4641
741 ----- 741
QY 4642 GAAGCGACAAAAAAGATTCGATAGACACACATTTGTTAAATGTTTACAGACTATAT 4701
742 -----AlaAlaGlyAlaGlyAlaThrAlaThrCysCysThrThr----- 753

Db 1022 SerAsnValThrPheAsnGlyHisLeuThrGlyThrLysLeuAsnGluPheAlaSerArg 1041
QY 6904 CCCCAATTTTGGAGATGGCAGAAATGGGCAAGAAATTTTGTATATAGAGAAAGAA 6963
Db 1042 ProSerPheLeuArgThrPheThrGluThrGlyAspGlnPheCysArgGluThr 1061
QY 6964 CAGTTGTTAAATTTGAGCGGGCTGTAAAGAAATATGAGTATAGT-----AGTAAT 7017
Db 1062 GlnLeuGlnIleLeuLysGluArgCysMetValTyrGlnThrAsnGlyAspLysGlyLys 1081
QY 7018 GAGGTTAAGACACAGAAATGTGACAGCGCTGTATACATATCAAAATTTTATTAAGAG 7077
Db 1082 AspAspLysLysGluLysCysThrGluAlaCysThrTyrTyrLysGluThrPheThrAsn 1101
QY 7078 TGGAAAATCGATATATGTAAGAAACAAAGAAAGTCAAAAAGTAAAGATGCAAAAAG 7137
Db 1102 TyrGlnAspAsnThrLysLysGlnAsnGlnArgTyrThrGluValLysGlyThrSerPro 1121
QY 7138 TATTAAGATTATCTCTTACTGAAAGACATAGACAGCAACATGTGCTCATATAT 7197
Db 1122 TyrLysGlu-----AspSerAspValLysGluSerLysTyrAlaHisGlyTyr 1137
QY 7198 TTTAACTGAAATTTAAAGAA---TATCTGCCAATPAGAT-----TGTCT 7242
Db 1138 LeuArgLysIleLeuLysAsnIleIleCysThrSerGlyThrAspIleAlaTyrCysAsn 1157
QY 7243 TGTATGCAAAAACCTTCTTACACAACTACCAAAAACACACACAAATATCCGAT 7302
Db 1158 CysMetGlu-----GlyThrSerThrAspSerSerAsn 1169
QY 7303 GCTAATGATATGCGCAGAAATGCGATATATTCCTGAAAGAAATTTAAACAG--TGTGAG 7359
Db 1170 AsnAspAsnIleProGlnSerLeuLysTyrProProIleGluGluGlyCysThr 1189
QY 7360 TGTCCGAACTTTCAAAAAGGATCTATGATTCATACAAAATTTACTGAACTTAA 7419
Db 1190 CysLysAspProSer---ProGlyGluValIleProGluLysLysValProGluProLys 1208
QY 7420 ATACTATGAAATTTGTAGAGAAAGCAGATTTATTTATCTAAAGAACAGAAATTAAT 7479
Db 1209 Val----- 1209
QY 7480 ATGATATTTACCTTGAGAAAGAAATTTATACATTTGAGTCTACAAAGAAAGAAAGT 7539
Db 1209 ----- 1209
QY 7540 AAAAATAGTTGACATATATATATCTTGCGATCTTAAGAAACCTTATGCACTGATAA 7599
Db 1209 ----- 1209
QY 7600 TATATAGAGAGAAAGAAACCTTGTGAAAATAGAGAAAGAAATCGTTTAAAGATTAAT 7659
Db 1209 ----- 1209
QY 7660 GAATGGAATGTTACAAAATTCAAAGTCTATACAGAGAGAAAGAGATGTGTACTCT 7719
Db 1209 ----- 1209
QY 7720 CCAAGAGAGACATATATGCTTAAAGAAATTTAGATGAATTTAAATTTGAAGACITTAAG 7779
Db 1209 ----- 1209
QY 7780 GATAGTATATATCTCTAAAATATGTTCTGTGAATCTACAGAAATGAAGAAATAGACATA 7839
Db 1209 ----- 1209
QY 7840 ATAAAAAATTTCAACTCAGAGAAAGGGTGGCAATGAAATCATATGTGATCTATGCAAA 7899
Db 1209 ----- 1209
QY 7900 TATAGTTTCGCTGATCTGGGTGACATAGTTAGAGAAAGATATGTACCAATTTGGTGT 7959

Db 1209 ----- 1209
QY 7960 TACTTACCTCCGTAGAAATTAATATATATAGSTTTTGAATACATATATGAAAAATGG 8019
Db 1209 ----- 1209
QY 8020 AGAAATTAATAAAGTAGAAATTAATAACMACAGATGTACAAACGTTCTGCTGTGG 8079
Db 1209 ----- 1209
QY 8080 TGGGATGCTAATAGAAAAGATATTTGAAAAGCAATGACGTGCAAAACACCAGAGATGCA 8139
Db 1209 ----- 1209
QY 8140 AAACTTTATAGAAAAGAGAAATGATGATTTGAACGCAATACATTAATACAGATTAAG 8199
Db 1209 ----- 1209
QY 8200 TGTGACATTAAGAGCAGATCCACCTGTGATGATTAATACCTCAACGTTTCGATGAGAG 8259
Db 1209 ----- 1209
QY 8260 ACTGAATGCTGATATATTTATTTAAAGCAGCTGATGGAAGAAATTTGAAAAA 8319
Db 1209 ----- 1209
QY 8320 TCATGTGATCAGCTGTAAACATCTGACAGATGCAAGAAATGATTAATGAAATAAGTGT 8379
Db 1209 ----- 1209
QY 8380 GAACAGTGTAAAGAGAGATGTCAGAAATTAATAAATTTGTCTTAATGAAATCTCTAA 8439
Db 1209 ----- 1209
QY 8440 TTGATATATACATCAATTAATATCAAAAGATTTATGAAACACCAATATATACAAAAATC 8499
Db 1209 ----- 1209
QY 8500 TCTACTTATGATCATGTCTCAAAATTTTGTACAAAAGTTGAAAACCTTTAAAGTGAATGT 8559
Db 1209 ----- 1209
QY 8560 TCTGTTAGAGCTTTCTGCAATCTTCATGAAACAAAGTAAGTGTGAATTAATTAATTT 8619
Db 1209 ----- 1209
QY 8620 AATGAAATGATGCTTCTTCCATATACGAAACATATGCTTTGAAAGAACACCAAAAAGT 8679
Db 1209 ----- 1209
QY 8680 TATTAAGAAAGCTTGACATTTGACACTACCTCTTAAGAAATCCATTTGATTAATGCTTACC 8739
Db 1210 -----LeuProLys----- 1212
QY 8740 GATCAAAAACAAAGATGAGTGAAGAAATTAACAAACCTTTACCTTCGTCGAGAAATGAT 8799
Db 1212 ----- 1212
QY 8800 TATGATATTAATCTTGATATATTTGAAACGCAATACCTTGTCTTAATAGTTCAATGATTAAC 8859
Db 1212 ----- 1212
QY 8860 AAAGTGTATTTGATTCCTCCAGAAAGAAAGACATTTATGTCAAGACCTATACACTGCATAT 8919
Db 1213 -----ProProLys----- 1215
QY 8920 AATTAATAGAAAAGGTGATTAAGAAATTTTAAAAAAAACCTTCTTACTCTGCTTTCAGT 8979
Db 1215 ----- 1215
QY 8980 CAAGCAATTTGTTAGTCAAAAATATTAATCGGAAGAAAGATGTGCTTTGAGGCAATG 9039
Db 1215 ----- 1215

QY	10111	TGGGAGCTATCTCTCGTGAATTACT--TCATCCGAAAGTAGATGATGAAGATTGGAT	10167
Db	1319	-----SerAspIleThrSerSerSerGluSerGlyArgGluLeuAsp	1333
QY	10168	ATTATGATATATATGTACACAGTACCTCAATATATAAAACATTGATAGAGATGACGA	10222
Db	1334	IleAsnAspIleTyrAlaProArgAlaProLysTyrIleThrIleuIleGluValValLeu	1353
QY	10228	GAACCAACCAAAAAAGCGATATACCAAGT-----GATGATACACCAAGT	10268
Db	1354	GluProSerGlyAsnSerGlnThrAlaSerGlyAsnAsnThrProSerAspThrGlnAsn	1373
QY	10270	-----AATGATACACGACGACGATACGAATAGATTATATGATGATGAATGAATGACATG	10320
Db	1374	AspIleGlnAsnAspGlyIleProSerSerLysIleThrAspAsnMetIleThrIleu	1393
QY	10321	AAACATGATTTTGTATCTCATATTTTACCAAAATACAGAACCAAT-----AATAT	10371
Db	1394	LysAspGluThrIleSerGlnTyrLeuGlnSerGluGlnProAsnAspValProAsnAsp	1413
QY	10372	TACAAAAGTCACAAATATTCACATGATACAGAACCTAATCTTATATCTGATATCT	10431
Db	1414	TyrSerSerGlyAspIleProLeuAsnThrGlnProAsnThrLeuTyrPheAspAsnPro	1433
QY	10432	GAAGAAAAACCTTTATATATCTATTCATGATAGGATTATATACGGGAAAGAAATT	10491
Db	1434	AspGluLysProPheIleThrSerIleHisAspArgAspLeuTyrSerGlyGluGluTyr	1453
QY	10492	AGTTATATATTAATATATGATGACTAATCTATATATGATATTCATGATAGCTAGAAAT	10551
Db	1454	SerTyrAsnValAsnMet--ValAsnThrAsnAsnAspIleProIleSerGlyLysAsn	1472
QY	10552	GATCTTATAGAGGATACATTAATTAATGATTCATGCTAGTGGTCTAAACCTATTGATA	10611
Db	1473	GlyThrTyrSerGlyIleAspIleuIleAsnAspSerLeu--AsnSerAsnAsnValAspI	1492
QY	10612	TATATGATCAACTATTG 10628	
Db	1492	IleTyrAspGluValLeu 1497	
RESULT 14			
AAR70235			
ID	AAR70235	standard; Protein: 700 AA.	
XX			
AC	AAR70235;		
XX			
DT	22-SEP-1995	(first entry)	
XX			
DE	P. falciparum EBL-e2.		
XX			
KW	Erythrocyte binding ligand; EBL-e2; binding domain; malaria; therapy;		
XX	vaccine.		
OS	Plasmodium falciparum.		
XX			
PN	WO9507353-A.		
XX			
PD	16-MAR-1995.		
XX			
PF	07-SEP-1994;	94MO-US10230.	
XX			
PR	10-SEP-1993;	93US-0119677.	
XX			
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.		
PI	Chituls C, Miller LH, Peterson DS, Sim KL, Su X;		
PI	Wellens TE;		
XX			
DR	WPI; 1995-123427/16.		
DR	N-PDB; AAG83528.		
XX			
PT	New erythrocyte binding domain polypeptide(s) - isolated from		
PT	Plasmodium binding proteins, used in diagnosis, treatment and		

PT prevention of malaria
 XX Disclousure; Page 56-57; 81pp; English.
 XX
 CC Erythrocyte binding ligand (EBL) family genes were cloned from
 CC P. falciparum chromosome 7 subsegment libraries constructed during
 CC genetic studies of the chloroquine resistance locus. The 4 genes,
 CC EBL-e1 (AA083526), E31a (AA083527), EBL-e2 (AA083528) and Pro13
 CC (AA083529), encode the proteins given in AAR70233-36, respectively. The
 CC binding domains of such proteins can be expressed e.g. in E. coli.
 CC Yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
 CC cells, and provide protection against P. falciparum.
 XX
 SQ Sequence 700 AA:

Alignment Scores:
 Pred. No.: 3.23e-64 Length: 700
 Score: 1032.50 Matches: 247
 Percent Similarity: 50.288 Conservative: 110
 Local Similarity: 34.798 Mismatches: 250
 Indels: 103
 Gaps: 22
 Gaps: 16

US-10-087-013-1 (1-10628) x AAR70235 (1-700)

QY 406 AATGATGCTATGACCTGTCGCCACCTAGAGAGACATATGTGTATAAACTTGGA 465
 DB 5 AAslyeValglYalacysAlaprotYargAlgluHlslencysAspYrAsnleuGl 24
 QY 466 GCCTCTAAATGATTAATACCCAAATATTCATGATTTATGGAAATGTACTAGTTACA 525
 DB 25 SerilleAsp-----ThrlThrSerThrThrHlslencysleuGlValCysMeAla 42
 QY 526 GCAAAATACAGAGATGATATGTTATTAATATCAT-----CCACATAAAGCACT--- 576
 DB 43 AlslsYtrclulglYasnsertlle---AsnThrHlslYtrHrclnHlslnslatYThrAsn 61
 QY 577 -----TCAGACCTTGTACTCTCTCTGTCACAGAGTTTGCAGATATAGTGAT 624
 DB 62 GluAspSerAlaSerGlInleucysThrValleuAlaArgSerPheAlaAspIleGlYAsp 81
 QY 625 ATTGTAGAGAAATGATATGATTT-----AAACCAATGTCCATGAC 666
 DB 82 IleValArglelYlYasnspleuYrleuGlYrAspAsnLysSluYsglInlArgLys 101
 QY 667 AAGATGAAAGCGGCTCCGAGAGGTTTTCAGAAATATACATGATGGAATGAGATGAA 726
 DB 102 LysleuGlInlYslencysAspIlePheLyslelHlsls-----LysAspVal 118
 DB 727 GTAAAAAATGAT-----TACATCTGATGCA---TCTGGAATATTAAT 768
 DB 119 MetlYtrHrAsnGlYalaglIngluArgYrIleAspAspAlaLysglYlYAspPhePhe 138
 QY 769 AATTAAGAGACAGATGCGATGATGGAATGAATTAATGAAGATGATTAATGAT 828
 DB 139 GlInleuArglelYlYasnspleuYrleuGlYrAspAsnLysSluYsglInlArgLys 158
 QY 829 GATGATCATATTAATCTGATATTTATGATGATGGAATGATGATGATGATGAT 888
 DB 159 HlslsAlaProlYsglInlYasnspleuYrleuGlYrAspAsnLysSluYsglYThr 178
 QY 889 AATCTTAATGCGGCCATAAAGAGAAAGGTTTCTACCAATTTAGATTGTGCTCCATCA 948
 DB 179 AsnGlYlncysHlslsYlleglYlYAspValProlYrHrAspYrValProlGl 198
 QY 949 TATTTACGTTGCTTCGAGCAAGAGGATTTTCCGAAAAAGAAATATTAATG 1008
 DB 199 lYrleuAlrYrHrPheglInlYrPalaGlInlAspPheCysArYlYsglYslYslYsl 218
 QY 1009 AAAAGATCAAGACTCTGCTGTAATGACAAAGAGCTTATATTTACTATCAATGAA 1068
 DB 219 GluAsnleuGlInlYsglInlYasnspleuYrleuGlInlAsnleuYrCysSerGlYAsnGlY 238

QY 1069 CATGATTTGACAGCAACTATTTGGAAAAAGATTTTTCGATTTGGATATAATAGTACT 1128
 DB 239 TyrAspCysThrLysThrIleYrleYslYslYslYslYslYslYslYslYslYslYsl 258
 QY 1129 GATGTTGCTCAATTAAGCAAGTTTGAAGTTTGGATTTGAGGATTAACAGACATTT 1188
 DB 259 AsnCysSerValrPrcysArYrMetYrGlYrHrPrlleAspAsnInlYslYslYsl 278
 QY 1189 AAAAACAACAAAAAATATGAAAAAATTA----- 1221
 DB 279 LeuLysGlInlYasnspleuYrleuGlYrHrPrlleAspAsnInlYslYslYsl 298
 QY 1222 -----CAATCATATTTATGCAAGCAATTAATTTGCAAT 1257
 DB 299 LysArgThrLysArgAlaAlaArgSerSerSerSerAspAspAsn----- 314
 QY 1258 AATATTAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1317
 DB 315 -----GlyYrGlSerLysPheYrLysLysLysLysLysLysLysLysLysLys 330
 QY 1318 ACTAATGACACTTTTAAATTTACTAATTAATGAAGAAAGATTTGTAAGGA----- 1368
 DB 331 AspValAspLysPheleuYslleuAsnLysglInglYllecYsglInlYsglInlProGl 350
 QY 1369 ---GGATTACAGAGAAAAAGATTTACTTTTACTATACAGTCTGATGCAAAAGGATA 1425
 DB 351 ValGlYAsnGlYlYslYslAspAsnValAspPheThrAsnGlYslYrYrYrYrYrYr 369
 QY 1426 TTTTATGCTGACAAATATGTCAGAGTGTCCGAGTGTGCGGGGTCAAAATGTATGAT 1485
 DB 370 PheSerArgThrGlInllecYsluProCysProlYrCysglYleuInlYsglYlPro 389
 QY 1486 AATATACACACACAAATACATGATGATGTCGAGTGTAAATGATGAAGCTTAACCT 1545
 DB 390 ProlYrPylValLysglYslYslYslYslYslYslYslYslYslYslYslYslYsl 409
 QY 1546 CCAATGGGTGTGAAGCTTAAATATGATGATGATGATGATGATGATGATGATGAT 1605
 DB 410 Lys-----AsnIleThrAspIleProlValleuYrProlYrProlYrProlYr 426
 QY 1606 ATTACACAAAAATTAAGAAATTTTGTACAGCTCAACTAATTAACAAAGTAAATTAAT 1665
 DB 427 IleuLysLysYr 445
 QY 1666 CAAAAATGGGATGCTATTAATGAGAT-----GAAATATTAATGATGAT 1710
 DB 446 LysLysTrpInlYslYr 465
 QY 1711 AATCTGGAACAAATCTGAAATCAATGATGATGATGATGATGATGATGATGATGAT 1770
 DB 466 ValGlInglYrHrPrcysLysPheThrGlInglYlYsglInlYrValLysSerYrAsnVal 485
 QY 1771 TTTTGAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1830
 DB 486 PhePheTrpAspPrcysAlaHlslAspMetLeuHlslAspSerValGlYrPrcysThr 505
 QY 1831 AAACTGTATTAATTAACACACAG-----CATGTATGATGATGATGATGAT 1878
 DB 506 SerLysCysLleAsnAsnAsnThrAsnGlYAsnThrCysArgAsnAsnAsnYslYsl 525
 QY 1879 AGAAATGCTTATGTTTGCAGATGGGTTAAACAAAGAGAGATGATGATGATGAT 1938
 DB 526 ThrAspCysglYslYslYslYslYslYslYslYslYslYslYslYslYsl 545
 QY 1939 AAGAACTGTTCACAAAAAAGATTAATGATGATGATGATGATGATGATGATGATGAT 1998
 DB 546 LysAspHisPheGlYslYslYslYslYslYslYslYslYslYslYslYslYsl 560
 QY 1999 CTTTGAAGGATTTATTTTAAAGTATGATGATGATGATGATGATGATGATGATGAT 2058
 DB 561 LeuIleValPheSerProlYrGlYlYslYslYslYslYslYslYslYslYslYsl 576
 QY 2059 AAGAACTTATGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2118


```

Db 577 GlysAlaLeuLeuGlnAlaHisLeuSAspValHisGlyAspThrAspAspIleuHisHisIle 596
Qy 2119 AGGAGCTATTAGAG-----AAT 2136
Db 597 LysLysLeuLeuAspGlnLysAlaValAlaValLeuGlyGlyLysAspAsnThr 616
Qy 2137 GCATATGAACTCTTGTATGATCATTAAGAACTGCCACGATATGTAAGACATATAT 2196
Db 617 ThrIleAspLysLeuLeuGlnHisGlnLysGlnAlaGlnLysGlnLysGlnLys--- 635
Qy 2197 ACAACGAAAGCATGTGA-----ACATCCCATATATCAACACCA 2235
Db 636 ---GlnGlnGlnLysGlnLysLysAlaGlnGlnGlnLysSerArgGlyArgSerLeuThr 654
Qy 2236 AACCCGTGTATTAACCTGTCGAGGACGACCAACCCATTAATATATAAGAAATAGCA 2295
Db 655 ArgGluAspGluArgThrGln-----GlnProAlaAspSerAlaGlyGluValGln 671
2296 CAATACCTTAAAGAGTGCATACGAGGAA 2325
:::
672 GlnGlnGlnAspAspAspAspTyrAspGln 681

```

RESULT 15

AAW22481 ID AAW22481 standard; Protein: 700 AA.

AAW22481; AC

07-OCF-1997 (first entry) DT

Plasmodium ebl-2. DE

DBL gene family; SAMP; sialic acid binding protein; vaccine; therapy; KM Duffy binding like gene; Duffy antigen binding protein; erythrocyte; DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response; Plasmodium.

Plasmodium falciparum. OS

MO9640766-A2. PN

19-DEC-1996. PD

07-JUN-1996; 96WO-US09508. PF

07-JUN-1995; 95US-0487826. PR

(USSH) US DEPT HEALTH & HUMAN SERVICES. XX

Chitlens C, Miller LH, Peterson DS, Sim KL, Su X; PI Wellens TE;

WPI: 1997-052231/05. DR N-PSDB; AAT72896.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC

This sequence represents ebl-2 of Plasmodium. Ebl-2 belongs to the Duffy binding like (DBL) family of genes which have homology to the Duffy antigen binding protein (DABP) and sialic acid binding protein (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The var family of genes modulate cytoadherence and antigenic variation of Plasmodium infected erythrocytes. SAMP and the Duffy antigen binding protein (DABP) are soluble proteins that appear in the culture supernatant after infected erythrocytes release merozoites. DABP and SAMP mediate the binding of merozoites and schizonts to the erythrocyte surface. These proteins are necessary for erythrocyte invasion by the parasite. This sequence can be used in the compositions of the

CC invention. The compositions are for the treatment and prevention of CC malaria, and comprise either a nucleotide sequence or encoded polypeptide CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a CC family of genes having homology with conserved regions of DABP and SAMP. CC The compositions are used for the treatment and prevention of malaria. CC They are also used in the preparation of vaccines for inducing a CC protective immune response in a mammal to Plasmodium merozoites CC (especially Plasmodium falciparum or Plasmodium vivax).

SO Sequence 700 AA;

Alignment Scores:

Pred. No.:	3,236-64	Length:	700
Score:	1032.50	Matches:	247
Percent Similarity:	50.28%	Conservative:	110
Best Local Similarity:	34.79%	Mismatches:	250
Query Match:	5.31%	Indels:	103
DB:	18	Gaps:	22

US-10-087-013-1 (1-10628) x AAW22481 (1-700)

```

Qy 406 AATGATGCTATAGCCTGTGGCCACCTAGAGACGACATATGTGTAATAAAGTGGAA 465
Db 5 AsnLysValGlyAlaCysAlaProTyrArgArgLeuHisLeuCysAspTyrAsnLeuGln 24
Qy 466 GCTCTAAATGATATTAATACCAAAATATTCATGATTTATGGAAATGTAAGTACAGTACA 525
Db 25 SerIleAsp-----ThrThrSerThrHisLysLeuLeuGlnValCysMetAla 42
Qy 526 GCATAATACGAAAGTGAATCATATGTTATATATCAT-----CCACATTAAGAACT--- 576
Db 43 AlaLysTyrGlnGlnLysAsnSerIle---AsnThrHisTyrThrGlnHisGlnArgThrAsn 61
Qy 577 -----TCAGACGCTGTACTGCTGTGACGACAGTTTGGACATATGATATGATGAT 624
Db 62 GluAspSerAlaSerGlnLeuCysThrValLeuAlaArgSerPheAlaAspIleGlyAsp 81
Qy 625 ATGTAGAGGAAATATATATGTT-----AAACCAATGTCCATGAC 666
Db 82 IleValArgGlyLysAspLeuTyrLeuGlyTyrAspAsnLysGlnGlnArgLys 101
Qy 667 AAAGTGAAGACGGGTCTCCGAGAGGTTTCAAGAAATACATGATGATGAAAGATGAA 726
Db 102 LysLeuGlnGlnLysLeuLysAspIlePheLysLysIleHis-----LysAspVal 118
Qy 727 GTAAAAAATGAT-----TCAATCCTGATGGA---TTCGAAATTTATAT 768
Db 119 MetLysThrAsnGlyAlaGlnGlnLysArgTyrIleAspAspAlaLysGlyLysAspPhe 138
Qy 769 AAATTAAGAGAGCATGTGGATGTGAATGAATAAATAGTATGGAAGCATTAACATGT 828
Db 139 GlnLeuArgGlnAspTyrPThrSerAsnArgGlnThrValTTrpLysAlaLeuIleCys 158
Qy 829 GATGATCATATTAATCGATATTTATGCAATACAGAAATATATACCATATTTTCA 888
Db 159 HisAlaProLysGlnAlaAsnTyrPheIleLysThrAlaCysAsnValGlyLysGlyThr 178
Qy 889 AATCTTAATATCGCCCATTAACAAGGAAAGGCTTCCATTAAGATTAATGATCCCTCAA 948
Db 179 AsnGlyGlnCysHisCysIleGlyGlyAspValProThrTyrPheAspTyrValProGln 198
Qy 1009 AAAAAGTCAAGACACTCTCTGCTGATGATACAAAGACCTATATGTAAGCATTAAGGA 1068
Db 219 GluAsnLeuGlnLysGlnCysArgAspTyrGlnGlnAsnLeuTyrCysSerGlyAsnGly 238
Qy 1069 CATGATGTAGACAACTATTTGAAAGAAAGCTATTTGCAATTTGGAATTAATAGTACT 1128
Db 239 TyrAspCysThrLysThrIleTyrLysLysGlyLysLeuValIleGlyGlnHisCysThr 258

```

QY 1129 GACGTTCGCAATAATGCAAGTTTGTGAGTTAGCAACAGACATTT 1188
 Db 259 AsnCySerValrPrpCysArgMetYrGluThrPileaspasnGlnLysGluPhe 278
 QY 1189 AAAAAACAAAAAGAAATATGAAAAAGAAAT- 1221
 Db 279 LeuLysLeuLysrArgLysYrGluThrGluLleSerGlyLysGlySerPro 298
 QY 1222 -----CAATCAATTTATGCAAGCATTAACAAATTTGCAAT 1257
 Db 299 LysArgThrLysrArgAlaAlaArgSerSerSerSeraspasn 314
 QY 1258 AATATTTATAGTGAATTTTAAACAATTTATGAAAAACTTAAGAAAGCAATATGCA 1317
 Db 315 -----GlyTyrGluSerLysPheTyrLysLysLeuLysGluValGlyTyrGln 330
 QY 1318 ACTAATGACACTTTTAAATTTACTTAATGAAAGAAAGATTTGTAAGCA- 1368
 Db 331 AspValAspLysPheLeuLysLleLeuAsnLysGluGlyLleCysGlnLysGlnProGln 350
 QY 1369 ---GATTTACACAGAGAAAAAGATATTACTTTACTAACAAGTGTGATGCAAAAGGATA 1425
 Db 351 ValGlyAsnGlnLysAlaAspAsnValAspPheThrAsnGlnLysTyrValLys---Thr 369
 QY 1426 TTTTATCGTTCAGAAATATTGCCAAGTGTGCCGACTGCCGGGTCAAATGTGATGATATA 1485
 Db 370 PheSerArgThrGluLleCysGluProCysProTyrProLysGlyLeuGlnLysGlyPro 389
 QY 1486 AATATACACACAAATCAGATATGATCGAAGTGTAAATATGATGAGACTTAAACCT 1545
 Db 390 ProTyrLysValLysGlnAspLysThrCysGlySerAlaLysThrLysThrTyrAspPro 409
 QY 1546 CCATGGGGTGTGAACCTACTAATATCATCTGCTTTATAGTGAATGCAACAGGTGAT 1605
 Db 410 Lys-----AsnIleThrAspIleProValLeuTyrProAspLysSerGlnAsn 426
 QY 1606 ATTACACAAAATATAGAAATTTTGTACACAGCTCAACTAATTAACAAGATAAAATAT 1665
 Db 427 IleLeuLysLysTyrLysAsnPheCysGlnLysGlyAlaPro---GlyLysGlnLle 445
 QY 1666 CAAAAATGGAATGCTTTTATAGAT-----GAAATATATATAGATGT 1710
 Db 446 LysLysTyrProLysTyrTyrAspGlnHisArgProSerSerLysAsnAsnAsnCys 465
 QY 1711 AAACGTGAAACAAATACGAATCAATATGATATATCTAAGATATATCATTTCAATAT 1770
 Db 466 ValGluLysThrTyrAspLysPheThrGlnGlyLysGlnThrValLysSerTyrAsnVal 485
 QY 1771 TTTTGAATTTAGGTTTACATATTTATTAAGGATACTATTAAGTGAATGACAAACTT 1830
 Db 486 PhePheTyrAspTyrValHisAspMetLeuHisAspSerValGluTyrLysThrGlnLeu 505
 QY 1831 AAACCTGTATATATATACACACAG-----CATGTATGATGAAATGTAC 1878
 Db 506 SerLysCysLleAsnAsnAsnThrAsnGlyAsnThrCysArgAsnAsnLysCysLys 525
 QY 1879 AGAAATGCTTATGTTTGCACAGATGGGTTAAACAAAAAGAAAGAAATAGTATATA 1938
 Db 526 ThrAspCysGlySerPheGlnLysTyrValGlnLysLysGlnGlnLysTyrMetAlaIle 545
 QY 1939 AAGAACTGTTCACAAAAAAGAAATATACAGCAATCGTATTTAGTAAATATATATAT 1998
 Db 546 LysAspHisPheGlyLysGlnThrAspIleValGlnGln-----LysGly 560
 QY 1999 CTTTGTGAAGTTATTTTAAAGTATGATGATTAACCTGACAAAGATGAAGCAAAATGG 2058
 Db 561 LeuIleValPheSerProTyrGlyValLeuAspLeuVal-----LeuLysGly 576
 QY 2059 AAAGAACTTATGAAAAATATAAAAAGAAAAAATGATTTTCCAAATTTGAAATATAT 2118
 Db 577 GlyAsnLeuLeuGlnAsnLleLysAspValHisGlyAspThrAspAspIleLysHisIle 596
 QY 2119 AGGACTATTTAGAG-----AAT 2136

Db 597 LysLysLeuLeuAspGlnGluAspAlaValAlaValLleGlnGlyLysAspAsnThr 616
 QY 2137 GCATAGAACTCTTGTAGTACACTTAAAGAAACTGCCACGATATGTAAGCAATATAT 2196
 Db 617 ThrIleAspLysLeuLeuGlnHisGlnLysGlnGlnAlaGlnGlnLysGlnLys---- 635
 QY 2197 ACAACGAGACATGTGAA-----ACATCCCATTAATGCAACACACA 2235
 Db 636 ---GlnGlnGlnCysGlnLysLysAlaGlnGlnLysSerArgGlyArgSerAlaGlnThr 654
 QY 2236 AACCGTGTGTAACTGTGTGAGGACGCCAACCCATTAATTAAGAAATATAGCA 2295
 Db 655 ArgGlnAspGlnArgThrGln-----GlnProAlaAspSerAlaGlyGlnValGln 671
 QY 2296 CAATCTTTAAAGAGATGCATACAGAGAA 2325
 Db 672 GlnGlnGlnAspAspAspAspTyrAspGln 681

Search completed: April 28, 2003, 11:29:39
 Job time : 1190 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: April 28, 2003, 10:25:46 ; Search time 34.0331 Seconds

(without alignments)
4316.649 Million cell updates/sec

Title: US-10-087-013-2

Perfect score: 19407

Sequence: 1 MGFSCKYFIKMGNAASLE.....IDLINDSLVILNLLIYMKY 3542

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

al number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	617	3.2	1435	EBAL_PLAFC	P19214 plasmodium
2	435	2.2	2869	RBP1_PLAFC	P006798 plasmodium
3	435	2.1	6669	NRBU_HUMAN	P20929 homo sapien
4	381	2.0	3135	S330_PLAFC	P08372 plasmodium
5	357	1.8	3210	CENF_HUMAN	P49454 homo sapien
6	342.5	1.8	1070	PYDR_PLAFC	P22290 plasmodium
7	337.5	1.7	1153	PVDB_PLAFC	P50493 plasmodium
8	332.5	1.7	1875	MRP1_YEAST	P00455 saccharomyc
9	327	1.7	1070	PVDG_PLAFC	P50494 plasmodium
10	326	1.7	1073	PVDA_PLAFC	P22545 plasmodium
11	317	1.6	1391	MST2_DROH	P08696 drosophila
12	313.5	1.6	2116	MYSD_DICDI	P08799 dictyostell
13	309.5	1.6	2748	NM1_YEAST	P00402 saccharomyc
14	309	1.6	2704	BPAL_HUMAN	P00301 homo sapien
15	299	1.5	2492	ATRX_HUMAN	P46100 homo sapien
16	298.5	1.5	2663	CENF_HUMAN	P00224 homo sapien
17	289	1.5	1630	MSPI_PLAFC	P04932 plasmodium
18	289	1.5	1639	MSPI_PLAFC	P04933 plasmodium
19	285.5	1.5	1957	YD86_SCHPO	Q10411 schizosacch
20	284.5	1.5	2476	ATRX_MOUSE	Q01687 mus musculu
21	274	1.4	2230	GOG4_HUMAN	Q13439 homo sapien
22	271.5	1.4	1790	USO1_YEAST	P23386 saccharomyc
23	270.5	1.4	3685	DMD_HUMAN	P11532 homo sapien
24	269.5	1.4	2022	ANT1_ONCOV	P21249 onchocerca
25	267	1.4	3678	DMD_MOUSE	P11531 mus musculu
26	261	1.3	1956	ATX1_PLAFC	P04956 plasmodium
27	259	1.3	3660	DMD_CHICK	P11533 gallus gall
28	259	1.3	5430	ACE7_HUMAN	Q09033 homo sapien
29	258.5	1.3	1638	YD67_YEAST	Q03661 saccharomyc
30	258	1.3	1679	YD67_YEAST	P40457 saccharomyc
31	258	1.3	2871	DESP_HUMAN	P15924 homo sapien
32	256.5	1.3	1727	ALM1_SCHPO	Q01687 mus musculu
33	256	1.3	1805	HMW2_MYCCE	P47460 mycoplasma

34	255.5	1.3	1557	DVAL_DICVI	Q24702 dictyocaulu
35	254	1.3	3680	DMD_CANFA	O97592 canis fam11
36	252.5	1.3	1928	MYSD_YEAST	P08964 saccharomyc
37	252	1.3	3418	BRC2_HUMAN	P51587 homo sapien
38	252	1.3	3911	AKA9_HUMAN	O99996 h a-kinase
39	249	1.3	5327	ACE7_MOUSE	O99420 mus musculu
40	245.5	1.3	1726	MSPI_PLAFC	P50495 plasmodium
41	243.5	1.3	1968	YD12_CAEEL	P34367 caenorhabdi
42	243.5	1.3	2339	RPC1_PLAFC	P27625 plasmodium
43	243	1.3	2349	TPR_HUMAN	P12270 homo sapien
44	242.5	1.2	1726	MSPI_PLAFC	P04934 plasmodium
45	237.5	1.2	1701	MSPI_PLAFC	P08569 plasmodium

ALIGNMENTS

RESULT	ID	EBAL_PLAFC	STANDARD	PRT	1435 AA.
AC	P19214				
DT	01-NOV-1990 (Rel. 16, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	01-FEB-1996 (Rel. 33, Last annotation update)				
DE	Erythrocyte-binding antigen EBA-175.				
OS	Plasmodium falciparum (Isolate Camp / Malaysia).				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxID=5835;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=90377299; PubMed=2204835;				
RA	Sim B.R.L.;				
RT	"Sequence conservation of a functional domain of erythrocyte binding antigen 175 in Plasmodium falciparum."				
RL	Mol. Biochem. Parasitol. 41:293-296(1990).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sdb.ch/announce/ or send an email to license@sdb.ch).				
CC	-----				
DR	EMBL; X52524; CAA36756.1; .				
DR	PIR; S11561; S11561.				
KW	Antigen.				
FT	DOMAIN 159 1104				
FT	ESSENTIAL FOR BINDING TO				
FT	ERYTHROCYTES.				
FT	VARIANT 1031 1031				
FT	E -> V (IN STRAINS FCR-3 AND ITG).				
SO	SEQUENCE 1435 AA; 167389 MW; 32A4309021B1C3D6 CRC64;				
Query Match	3.2%; Score 617; DB 1; Length 1435;				
Best Local Similarity	21.8%; Pred. No. 3.3e-21;				
Matches	332; Conservative 215; Mismatches 537; Indels 418; Gaps 69;				
QY	1101 EACNEYNIDIGLKEQ---WNIIIDRYKELHQAQMSVNSGIEASTAKNHIIDRVIEF				1157
DB					
QY	20 KARNEY-DI-----KENEKFLDYKKEKFNELDKKKKGNVQKT-----DKIIFEF				62
QY	1158 L-SELXQNGGSKNS-----GTSDESAVIGTWTYYENVCAIYLDHGTGDFDCDSQNEFCDE				1212
DB					
QY	63 IENKLDILNNSKFNKRWKSGTTPDN---IDKNMSLN-----				96
DB					
QY	1213 KSDGKNEKYPARPDQDDGACGSGSKPTFRVQ-IKTKKAEEKDECTKYNDILKEN				1271
DB					
QY	97 -----KHNEEMNNNTQSLSTSSLIKQKKYPIINAVRSILSFDSR---INNRRNTS				149
DB					
QY	1272 DGRKQVEDCHPKNSNGSDYDQGNINLVEDPRVCAPPRKRLCYHFLANDNEIKKLQSQ				1331
DB					
QY	150 SNNEVLSNCRERKRG---MKWCKKKKN-DRSNVVCIPDRIDLCIVNLS---IKITYRK				201
DB					
QY	1332 VMLKEAFITSAALAEFTFSYTYTKSKDGEQNEELDKELCKIPAPFLRSKFTYFGDYRDL				1391
DB					

```

Db 202 EYKDHFIETASKE---SOLLKKNNDKNYS-----KFNNDKNSEFLYGHLA 246
OY 1392 FGTIDSKGHGESSKLEQIDSLFKNDOKSPNGKT---ROEWTEHSHEIWEAMICALVY 1448
Db 247 MGNDDMOFG-GYSTKAMENKIQEYFKAGHGEISHKIKNFKEWNNERREKLMZAML----- 300
OY 1449 IAKKDDFLENTGYNNVKSDDKSTLLEFPKRPQLRWLTLEWYDYCYTRORYLADVOEK 1508
Db 301 -----SHKKNIN-NCKNIPDEEL-----OITOWIKENHGEFLERDRNSKLPKSK 345
OY 1509 CKNNDOLK-CDEECNKKCEDYKY-MKKKEWIPDKYKDKRDRKRDROHIGWAVDY 1566
Db 346 CANNLLEYKCEKECJDPCMKYKDMIIIRSEFENHTLSKETETOKVKE----- 392
OY 1567 TGTNATDYLNKFTASCCKGPGSASVORNIOLLEKQAYYADKQCC---TKFTE----- 1619
Db 393 ---NAENYL-IKISEKNND-----AKVSLLNCOAEKSKYCDCKHTTLYKSVLN 439
OY 1620 -ND---DKYTNISSKDKCK-GLVKEA-WTGAIKQKNGPNYNNLKELTEDYLFPSRRL 1672
Db 440 GMDNTIKERREHIDDDSKFGCDNSVDNTNWKVECKNPIYS-----TKDYCVPPRRQ 494
OY 1673 RICFHALDGNYPYDKENGEL--RRKLEAVATEGYNLGOYKKEKKEKIKTSDAHY 1730
Db 495 ELCLGNDRIRY-----DKNLMIKEHILAIYESRIKRYKKNKDKKE----- 538
OY 1731 STEVPCSAMKSYDLDIIGIDNLEDEKQTEENLKIPKNGTSVGKSGDSTGNP 1790
Db 539 -----VCKIINTFADIDIDIIIGTGYMND---LSNRKLYGKINTSKYVHRKKN----- 585
OY 1791 GSTARFEMNEKCEVNNAMICGYRGDNGSGNSASDEDLKCGSVSPDDYPMKGN 1850
Db 586 DKLPEDEMKYIKKQVMMVNIWVEF-----DKTVCKE 617
OY 1851 RDEGTAFOELRFAWGEDFCRKHKEKELEKLYGACNDYTCGDNEDKRRKCTDACYOYKF 1910
Db 618 DDIENIPOEFREFSEMGDYODCKMIELLVEKKEKPCDD-----NCKSKCSYKEM 672
OY 1911 ISEMKPOYKOKKYGK---NKDKIYSEHPYAKADEAREYLDKOLKICENKSGDCEY 1966
Db 673 ISKKEEYNKQAKOYOEYOKGNNYKMYSEFKSIKEVYLYKIKYSEK-----CSNLFEDF 727
OY 1967 K-----C--MKDYSTORLTGDSNOMPASIDDEPKEX-----ECKKNCQYVRG 2007
Db 728 KEELHSDYKNNCTMCPKPKVDPIK-ITNNNOTSEAVPEENTELAHNTEPPSISEGPK 786
OY 2008 PPRVARETPSPRVLISKATASKEAKTAPTKOPKVENLTTEMRQOTRRRAAQOTRK 2067
Db 787 NEOKERDDDS-----LSKISYS-----PENSRR-----ETDAK 814
OY 2068 RTSVATTESVVG-TMVKALISNKPDSNGGLEGNPKTYGQYPMGCIYKSGKENENGIC 2126
Db 815 DTSMLKLKGVADISMPKAVIGSSPN-----DNIN-----VIEOGNIGSV- 855
OY 2127 MPPRRKILCINNIYLYETENKRDNDIKAEFICAIAETQFLMKYIIEENPAENELON 2186
Db 856 -----NSRPLSDVAPDKKELEDONSDESEYV-----NHSISPSINNDDGS 899
OY 2187 GTIDPEFRILMYTYGYKDMFFGTDISNDKRIITVNSVTIINENNNKKOKDKKDEEL 2246
Db 900 GS-----GSATVSESSSNTGLSIDDRNGDTFVRTOTANTAEVYIREKNDKODE 951
OY 2247 RKIWEKNKKFIWGMATYGLYHITDEN-----EKEKRDNYOYDMTKLAPSLPEFYKRP 2302
Db 952 KGADEER-----HSTSELSPEKEMLTDEGNSLN-----HEEYKEH 990
OY 2303 QFLRMTEMAEEFCNKRREOLK-----LEAGKEVEGNSN----- 2339
Db 991 TSNNDNYOOSGIVYNNVVEKAKLTLENPSSSIDEGAHHELSEPNLSSQODMSNTPGPL 1050
OY 2340 DKGTOEACACVYQNFIKKWKTEYEROREKFKD-----KDGKKYKDYPSSTER 2388

```

```

Db 1051 DNTSEETTER--ISNNEXKVEREDEFTIYKEDYILVLSHNNRESDDGELYDE-----NS 1104
OY 2389 DIEKATCAHEYNIMTKELCGNCKOSQKPSQPKTTOOSQSSDANDAPESID-----Y 2444
Db 1105 DLSTYNDSSEDAEAKMK---GN-DTSEKSHNSOIHESDOCK-----NDMTYVGLGTTH 1155
OY 2445 VPEEFNKCEPELSK-----GSMIHTKITEPKIPNKCVEKAAAYLSKEAENNDITLK 2499
Db 1156 VONEISVPTGEIDELKLESKESKH--KAEERLSHDHDK-----INPEDRNSNTLTK 1209
OY 2500 EKFTIESTEKESKNSNTNN 2521
Db 1210 D-----IRNEEMERHLTON 1224

RESULT 2
RBPL_PLVAB
ID RBPL_PLVAB STANDARD: PRT: 2869 AA.
AC Q00798;
DT 01-Apr-1993 (Rel. 25, Created)
DT 01-Apr-1993 (Rel. 25, Last sequence update)
DT 01-Oct-1996 (Rel. 34, Last annotation update)
DE Reticulocyte binding protein 1 precursor.
GN RBPL.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; Pubmed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
  merozoites."
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
  HUMAN RETICULOCYTE CELLS.
CC -1- SUBUNIT: HOMODIMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation-
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb.ch/announce/
  CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M88097; AAA29743.1;
KW Malaria; Receptor; signal; Transmembrane.
FT SIGNAL 1
FT CHAIN 17
FT FT 18 2869 RETICULOCYTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205BECFF CRC64;

Query Match 2.2%; Score 435; DB 1; Length 2869;
Best local Similarity 18.0%; Pred. No. 2.2e-12; Indels 984; Gaps 129;
Matches 529; Conservative 444; Mismatches 978;

OY 14 MAASLEGDAKSPILKESHKSARNVLERYAKNIRH-PSKYA-KEHVDLSKGLDTKAEFRG 71
Db 428 DAAKLES-----INDEBSAEIDLOKFPDNKALANNVCIFQYIKTINEPIKRAY----- 478
OY 72 GPSTPYKHNHYTYPCNDLKHETMLKRDVNLRHPCGREGQNRDEDESECGNKIRN 131
Db 479 -BSKYIKSNELLSTIIDTLGKGTAL-----GESTPDQ---EC-NKIKT 518
OY 132 --YKRNDAIACAPPRRRHCKN-----LEALDINTONHDLGNVLTAKREGS 182
Db 519 EAEKVADDA-----EDICEKNEQIYETIPESEDETIDDKINDLODLIDOKREYKDE- 569

```

QY 183 IYNNHFKGTSACTALANSFADIGIVRGIDMFKPNVDKVTGLREVFKKIHDMGME 242
1111 : : : : :
Db 570 IYNNSEF-----ISNRKNIYE-----NLKETYEETELNDICK----- 601
QY 243 VAKNDYPPDSGNYKRLREAMNVRNKKWEAT--TCDASYS----- 282
1111 : : : : :
Db 602 LEND-----TSKVNFYLMQIRKINTEKRTIDESLQTVKEFKELIDSEKITEKEFEKSV 658
QY 283 ---GYPMOSESNTPLFSNPKCGHKQ-----GKPNLIDYVPOYLR-----WPMDSG 326
1111 : : : : :
Db 659 TEINRQDESARDL-----HEBQIKELIDMAKKVYIKELLSLKKSSYFFEMME 711
QY 327 ---EFCRKNITKLLKVVDSCHNDKERLYCSHNGDCTTTT-----KGLIHL- 371
1111 : : : : :
Db 712 LINTASYDMGEGSAAKKE--ADNDINALYNSYREDINALIEVEKPYTENKSTLEML 769
QY 372 -----DNKCTDSTFCRY----- 384
1111 : : : : :
Db 770 KDBEMEELQDAKETPAKLNFEVSDDKLTVDYTKMSAEVTNAGIKKELOKOPENVHKK 829
1111 : : : : :
QY 385 -----FEVWLGNOO-----EAFKQKEKEKEIOSYLSNDNKFVNINSEYKQ 428
1111 : : : : :
Db 830 KERSDAFSTFKFALQNSMOYNGSDALIEKHQNRSEKEBEYFKNSVEEDLSREETEBQ 889
QY 429 FYEKLK-----ETO---YATNDFLNL-----NEGKYCKGL 458
1111 : : : : :
Db 890 EYFKHKNFSRRKGEISAEITNMREVINKIESOLANYGYEKFYSLIDQNEVSTAK--- 946
QY 459 PGKDLFTFNNSADDKIEFRSEY-----CQVCPGOCVGDIXTHKSDND--- 504
1111 : : : : :
Db 947 -ALKEIVSOSLRDKIDQYETEKETSVAVENTVSTIOSLSAIDSLKRLNGSINNCKY 1005
QY 505 -----RRVNNEDYK--PBGVKEPTNITVLVSGNEQDITOKLEPNFNSST 548
1111 : : : : :
Db 1006 NNDIDILRSKITLREYQKEMPKRGDKGENTALLKSLRKMCKINELKNDGLNSL 1065
QY 549 NYKDKANNQWECY-----KDBN-----INCKLEONTNINNDPKT--ISHNFF 552
1111 : : : : :
Db 1066 DPKREDELK--FYSESKSRIHLSDOKGPODPLNR--IDEMEDIKRDVDELVANVOVIS 1120
QY 593 ELWVYTLIDTLKWNKLTCTCINNTTHCIDECNRNCLCFDRVYQKKEEWMNSIKLFTK 652
1111 : : : : :
Db 1121 ENKVTLEFKNSVYIEAMHSHT--NTVAHGTS--NKNEIL--KSVKEVEDLNV--- 1171
QY 653 KKNIOOSYSYNIN-----LEFGYFVKYMDKLDKND-----BAKMELENIKRKK 697
1111 : : : : :
Db 1172 ---QNEDEYKVKVNPENKQLEAIRGSMKLEKVINRHVSEMTQLESTANTLSNKGRE 1227
QY 698 NEFSNLENNRDYLENAIELLDLAKETATYCKDNNTNEACETISHNATTNFCVPRGSTOP 757
1111 : : : : :
Db 1228 NEHDELELNR--TKGMRDIEYKIKIAELKGTVELELDANERANK--VBP---EP 1278
QY 758 TKRIKEIAQYEFKRSAYEARN-----RGLHKLKGAHGGIYKRGRRRDFDNCRIMI--K 812
1111 : : : : :
Db 1279 ERRI--IGHVLEIYIEKDKAGVDEANSLKRIKLELOETS---DSONELEYTTSITK 1333
QY 813 HSNRNLGFSNGPCDGTGDTGQIOTREVVGTWEVDEHMKKDEHYIMPRRRHICTSNL 872
1111 : : : : :
Db 1334 HLENAGY-----EDVIKRNEEDSIOREKAKSLLET 1365
QY 873 EHIQTDHPLNGIYVDDVYNNSLGDLVLSAKYEANKIIRMYEKNNKLPKRYTDPKQ 932
1111 : : : : :
Db 1366 DEKKKLVQVNNMLQSAIQNAGIS-----KELNELKVIILLSTNY 1408
QY 933 TITICRAIRSFADIGIIRGLDMERNGDVRLQGHLETFVGNHRSLSKGGNDKYDND 992
1111 : : : : :
Db 1409 SSI---LEYKKNSSSEYFESOL--ANGETKAGE----- 1439
QY 993 PKYLLRENNMEAN--RAKYWEAMKCDIKYLLKDSGHSQSTOSSYCYGSDHTPLDDYIYOKL 1051
1111 : : : : :
Db 1440 -----EKNASARLAEAEKLEKQIYKD-----LDYSDIDDKVY--- 1471

QY 1052 RMWTEAEMTCVQV--KKEYDKLEKCKECKDKDNQOGCTKESGTCTCTEACNEYNDI 1109
1111 : : : : :
Db 1472 -----KIEGIRKREILKKEESA----- 1487
QY 1110 IGLMKRQNNIISDKYKLEHQQAQMSVNSGIEASTAKNHIIDNVIPLSELYQONGKS 1169
1111 : : : : :
Db 1488 LTFMEZ-----SEKFRQM-----CSSHMEAKKGGKKE---YLRKNGDG 1524
QY 1170 NKSJSD-----ESAVIGTNTTYENVAGYIHDGTFDDQSQONER 1209
1111 : : : : :
Db 1525 GKANIDSQMEVGNVSAEHAHFHTVEAQVDTKAFCSIVAYVTRKMLNFHESLMKEV 1584
QY 1210 ---CDEKSDGKNEKAFRDKPODHDGACGCKSGSPTRY--QITKTKAKEKTEBEKTYN 1265
1111 : : : : :
Db 1585 KVCEKKNND--EAEKYSAKLR--YDGRKARVSENERKISELKERAKYEEK--ESSQLN 1638
QY 1266 DILKENDGKQVDECHPKRKNNGYPPMOCGINLVDPDPRCMPPROKLCVHLANDNEI 1325
1111 : : : : :
Db 1639 DV--STKSLQIDNCRQOLDS-----VLSNIGRV-----KONALQYFDSADKSA 1680
QY 1326 KILQ--SOVNLKBAFTKSAAEETFFSWYKSKDGSGNELDK--ELKEGKIPAPFLRSMFY 1382
1111 : : : : :
Db 1681 KSVLPISLGAEKSLDKVYKAKE-----SYEKNIETVQNMESRLNVEGSL----- 1726
QY 1383 TFGDYDFLEGTDISG---HEGSKLKQ-----IDSLFKNGDQSPNGKTRQEMWTE 1433
1111 : : : : :
Db 1727 -----TDIDKRTIDENDLLKMKROYEGGLQIKENADR---KSNFELVGS 1771
QY 1434 HSHIEWEAMCALVKIGAKKDDFT-----ENYG-----YNNVK-----FSDKSP 1472
1111 : : : : :
Db 1772 EINALDPPSTIFIKLKYDWTGDKNIGVAKNEIHGEFTSYNLTIEHLSNADYSV 1831
QY 1473 TLEEFARPOFLRMLETWYDDYCYTRQ-----KYLDVQEKCKSN--- 1512
1111 : : : : :
Db 1832 TFE-----KAQSLRELAKKEEHLRRREEAIFLLNDIKKYESLKLKEMKKYSAVEEGM 1887
QY 1513 -----DOLKC---DTECNKCEDYVYMKKKKEMIPODKYKDEROK 1351
1111 : : : : :
Db 1888 KRDTSVSOLVODMKTIVBELKTLNDISECSSVANNVVSIVKRYKE---SKHADYRDA 1943
QY 1552 KRFDRQHIGVWVDYGTNTDYLNRKFTASGDKPGSA-----SVQORNI 1597
1111 : : : : :
Db 1944 NSMTESNV-----TLNATFLSDEAKISSGMEFRAEKSNFKTDLEITFSVSN 1594
QY 1598 QLEKOAYYDADKHGCGTKEIENDDKYTNISSKDKGKGLYEANTGAIKQNRGNPNYNN 1657
1111 : : : : :
Db 1995 ELKK-----IEQDSNDV--IQKRESQQLKADP-----DIYN- 2026
QY 1658 LKELTEDEVLPFSRLRICFHALDGNITDPEYKDENGILRKLMEVAATEGYNLGOYKREK 1717
1111 : : : : :
Db 2027 -----VIKLNENEKLEEA-----KNKE 2045
QY 1718 EKEKIRTSDAKXYSEVPKSAKMYSPYDRLDILGDNJ-----DEKOTTEELRK 1770
1111 : : : : :
Db 2046 EYVSEKREALKLSQYEBGIRCFENFHLDDTELEENLKNWYTIYRDKKSERESLOE 2105
QY 1771 IFNKNGTSVGKSGDSTGTGNGSTARKFEMNKECVWMMAMICGYKRGDRDNGSNGSARSD 1830
1111 : : : : :
Db 2106 MENEMNTY---SNSITQLEGIVASA---DESKEDI-----EKLERSN 2141
QY 1831 EDLK-----KGSVPS-----DDDYPMGKRNDETAQVFLMFAWGEDCKHKEKE 1877
1111 : : : : :
Db 2142 EEMRNISEKSTIDSKVIEKNSSTIDELYKIGKRCQ-----AHWISLISYANMK 2190
QY 1878 LEKLVGACNDYTCGDNEDKRCKCTDACTOYKFRFLEWKPOYEKOIKKYGKNNKRIYSEH 1937
1111 : : : : :
Db 2191 TSKKILMIN-----KKEKNTKCVYIKXNSSSTQDYEVELKGFYSKILFSSASG 2241
QY 1938 VANDAE---DAREYLDKOLKIKCEKNSGDCXYKCMKDVSTORLTDGNSQNMMA---SLD 1990
1111 : : : : :
Db 2242 IYONADYVSVNPAKHESINAIYRDIKELLYLPHQNSDIS---IYEGVQNNMLALYDKLN 2298
QY 1991 DEPEVEGKNCQVPRGPPRYVRETSPRVSLSKATASKAKEKTAIAPTOPKAVENTLT 2050


```

Db 2299 EERKRD-----ELYNISSTKLOHSHSTVYK-----PMIEHKGMN--- 2337
OY 2051 EMRAQTRRAAQQTRKRTSTATTTESDVGTWKALISNPKDSRGIE----- 2098
Db 2338 -----ETNNKSLKEKKKSVNDHMSMEMEMKNGKLTPEVONINNTYSIEAEVKT 2393
OY 2099 -----GCNPTYGQYKRWGCIYVGS-----KENENGICMPRRKRLCINN 2138
Db 2394 LEEIDRDYGNQYVEEHKKQFSILIDRTVALMDIDIEIFKKN-----YN 2439
OY 2139 IOYLNTEERKRDNDKEAFKCAIETQFMKYIENPAAENELONGTIPDEFRIMT 2198
Db 2440 LMEVNETIHRVNDYEKTKNKLYVQKTE-----EQILENKDDMDLONIFKKSIIEX 2496
OY 2199 YTYGDKDMEFGTDISNDKKIITVNSVTI-----LNENKKKKDKKDELRKIFWEKKN 2255
Db 2497 FENVKKKKKSILNDIYEOEELKIGELDEIKRNVETLSYSYIDCKMEMSKNLEKKS 2556
OY 2256 KFIWEGMIYGLVHLLDEN-EKEKIDNYQYNDMTKLTSLFEV-KRPOFLMPTW-- 2311
Db 2557 KMANVTSTLEEREANEINRDAQIND-----DDTILNSVLEAALQKRGDMDAIFSQMSA 2611
OY 2312 -----AEFCNRKKDQLKLEAGCKEYECNGSNDKTOCAGCTYQNFIKKMT 2362
Db 2612 DNNPNEYKSAEKYMNANILIQLEVKLR-----GOLYQDSHIL-----S 2654
OY 2363 ETEROREKFKDQNGKRYDYSTERDIEKATCAHEYLNMKLELGNKDCSCMKPSSQ 2422
Db 2655 ENMSKKSALIEKERTARALTSNNRREBE-----RAVOEKMNNDP--QSETH 2704
OY 2423 LKPTQOOSQSANMPESLDVPEPEKNCCEPELSKGSMTHTKITEPKIPANCVEKA 2482
Db 2705 SCSGSGEGKESDSDEGLHDAGADSDTS-----SAKGA--HELEEEETTPM----- 2751
OY 2483 AYYSKREANNNDITL-----KEKPIESTKE-----KESKNSWTNNPCDPKPYAP 2531
Db 2752 -----EEPEMN-DNTLLGVDTJRSDPEMDHENTODGYTDISNSDEADILNGKFNPN 2805
OY 2532 DKYTG 2536
Db 2806 VYTAG 2810

```

```

RN [3]
RP STRUCTURE BY NMR OF 6610-6669.
RX MEDLINE-98179559; PubMed-9514727;
RA Pollou A.S., Millevol S., Gaudel M., Kolmerer B., Pastore A.;
RT SH3 in muscles: solution structure of the SH3 domain from nebulin.;
RL J. Mol. Biol. 276:189-202(1998).
CC -1- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MAINTAINING
CC THE STRUCTURAL INTEGRITY OF SARCOMERES AND THE MEMBRANE SYSTEM
CC ASSOCIATED WITH THE MYOFIBRILS. BIND AND STABILIZE F-ACTIN.
CC -1- TISSUE SPECIFICITY: MUSCLE-SPECIFIC. LOCATED IN THE THIN FILAMENT
CC OF STRIATED MUSCLE.
CC -1- DISEASE: Defects in NEB are a cause of the autosomal recessive
CC form of nemaline myopathy (NEM2).
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 178 NEBULIN REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X83957; CA58788.1; ALT_SEQ.
DR EMBL; M19668; AA59916.1; ALT_SEQ.
DR EMBL; M19669; AA59917.1; ALT_SEQ.
DR PIR; A29979; A29979.
DR PIR; B29979; B29979.
DR PDB; 1ARK; 28-JAN-96.
DR PDB; 1NEB; 24-DEC-97.
DR Genew; HGNC:7720; NEB.
DR MIM; 161650; -.
DR MIM; 256030; -.
DR InterPro; IPR000900; Nebulin.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00880; Nebulin; 146.
DR PRINTS; PR00510; NEBULIN.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00237; NEBU; 181.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW Actin-binding; Muscle protein; Cytoskeleton; Repeat; 3D-structure.
FT REPEAT 113 107 NEBULIN 1.
FT REPEAT 113 143 NEBULIN 2.
FT REPEAT 148 178 NEBULIN 3.
FT REPEAT 183 213 NEBULIN 4.
FT REPEAT 218 248 NEBULIN 5.
FT REPEAT 253 283 NEBULIN 6.
FT REPEAT 289 318 NEBULIN 7.
FT REPEAT 324 354 NEBULIN 8.
FT REPEAT 363 393 NEBULIN 9.
FT REPEAT 398 428 NEBULIN 10.
FT REPEAT 434 464 NEBULIN 11.
FT REPEAT 502 532 NEBULIN 12.
FT REPEAT 537 567 NEBULIN 13.
FT REPEAT 573 603 NEBULIN 14.
FT REPEAT 611 641 NEBULIN 15.
FT REPEAT 681 711 NEBULIN 16.
FT REPEAT 749 779 NEBULIN 17.
FT REPEAT 784 814 NEBULIN 18.
FT REPEAT 820 850 NEBULIN 19.
FT REPEAT 858 888 NEBULIN 20.
FT REPEAT 893 923 NEBULIN 21.
FT REPEAT 924 954 NEBULIN 22.
FT REPEAT 959 990 NEBULIN 23.
FT REPEAT 993 1023 NEBULIN 24.
FT REPEAT 1028 1058 NEBULIN 25.
FT REPEAT 1064 1094 NEBULIN 26.
FT REPEAT 1102 1132 NEBULIN 27.
FT REPEAT 1137 1167 NEBULIN 28.
FT REPEAT 1168 1198 NEBULIN 29.

```

FT	REPEAT	1204	1234	NEBULIN 30.
FT	REPEAT	1237	1267	NEBULIN 31.
FT	REPEAT	1272	1302	NEBULIN 32.
FT	REPEAT	1308	1338	NEBULIN 33.
FT	REPEAT	1346	1376	NEBULIN 34.
FT	REPEAT	1381	1411	NEBULIN 35.
FT	REPEAT	1412	1442	NEBULIN 36.
FT	REPEAT	1448	1478	NEBULIN 37.
FT	REPEAT	1481	1511	NEBULIN 38.
FT	REPEAT	1516	1546	NEBULIN 39.
FT	REPEAT	1552	1582	NEBULIN 40.
FT	REPEAT	1590	1620	NEBULIN 41.
FT	REPEAT	1625	1655	NEBULIN 42.
FT	REPEAT	1656	1686	NEBULIN 43.
FT	REPEAT	1692	1722	NEBULIN 44.
FT	REPEAT	1725	1755	NEBULIN 45.
FT	REPEAT	1760	1790	NEBULIN 46.
FT	REPEAT	1796	1826	NEBULIN 47.
FT	REPEAT	1834	1864	NEBULIN 48.
FT	REPEAT	1869	1899	NEBULIN 49.
FT	REPEAT	1900	1930	NEBULIN 50.
FT	REPEAT	1936	1966	NEBULIN 51.
FT	REPEAT	1969	1999	NEBULIN 52.
FT	REPEAT	2004	2034	NEBULIN 53.
FT	REPEAT	2040	2070	NEBULIN 54.
FT	REPEAT	2078	2108	NEBULIN 55.
FT	REPEAT	2113	2143	NEBULIN 56.
FT	REPEAT	2144	2174	NEBULIN 57.
FT	REPEAT	2180	2210	NEBULIN 58.
FT	REPEAT	2213	2243	NEBULIN 59.
FT	REPEAT	2248	2278	NEBULIN 60.
FT	REPEAT	2284	2314	NEBULIN 61.
FT	REPEAT	2322	2352	NEBULIN 62.
FT	REPEAT	2357	2387	NEBULIN 63.
FT	REPEAT	2388	2418	NEBULIN 64.
FT	REPEAT	2423	2453	NEBULIN 65.
FT	REPEAT	2456	2486	NEBULIN 66.
FT	REPEAT	2491	2521	NEBULIN 67.
FT	REPEAT	2527	2557	NEBULIN 68.
FT	REPEAT	2565	2595	NEBULIN 69.
FT	REPEAT	2600	2630	NEBULIN 70.
FT	REPEAT	2631	2661	NEBULIN 71.
FT	REPEAT	2666	2696	NEBULIN 72.
FT	REPEAT	2699	2729	NEBULIN 73.
FT	REPEAT	2734	2764	NEBULIN 74.
FT	REPEAT	2770	2800	NEBULIN 75.
FT	REPEAT	2808	2838	NEBULIN 76.
FT	REPEAT	2843	2873	NEBULIN 77.
FT	REPEAT	2874	2904	NEBULIN 78.
FT	REPEAT	2909	2939	NEBULIN 79.
FT	REPEAT	2942	2972	NEBULIN 80.
FT	REPEAT	2977	3007	NEBULIN 81.
FT	REPEAT	3013	3043	NEBULIN 82.
FT	REPEAT	3051	3081	NEBULIN 83.
FT	REPEAT	3086	3116	NEBULIN 84.
FT	REPEAT	3117	3147	NEBULIN 85.
FT	REPEAT	3152	3182	NEBULIN 86.
FT	REPEAT	3185	3215	NEBULIN 87.
FT	REPEAT	3220	3250	NEBULIN 88.
FT	REPEAT	3256	3286	NEBULIN 89.
FT	REPEAT	3294	3324	NEBULIN 90.
FT	REPEAT	3329	3359	NEBULIN 91.
FT	REPEAT	3360	3390	NEBULIN 92.
FT	REPEAT	3395	3425	NEBULIN 93.
FT	REPEAT	3428	3458	NEBULIN 94.
FT	REPEAT	3463	3493	NEBULIN 95.
FT	REPEAT	3499	3529	NEBULIN 96.
FT	REPEAT	3537	3567	NEBULIN 97.
FT	REPEAT	3572	3602	NEBULIN 98.
FT	REPEAT	3603	3633	NEBULIN 99.
FT	REPEAT	3638	3668	NEBULIN 100.
FT	REPEAT	3671	3701	NEBULIN 101.
FT	REPEAT	3706	3736	NEBULIN 102.

FT	REPEAT	3742	3772	NEBULIN 103.
FT	REPEAT	3780	3810	NEBULIN 104.
FT	REPEAT	3815	3845	NEBULIN 105.
FT	REPEAT	3846	3876	NEBULIN 106.
FT	REPEAT	3914	3944	NEBULIN 107.
FT	REPEAT	3949	3979	NEBULIN 108.
FT	REPEAT	3984	4014	NEBULIN 109.
FT	REPEAT	4021	4052	NEBULIN 110.
FT	REPEAT	4057	4087	NEBULIN 111.
FT	REPEAT	4088	4118	NEBULIN 112.
FT	REPEAT	4123	4153	NEBULIN 113.
FT	REPEAT	4156	4186	NEBULIN 114.
FT	REPEAT	4191	4220	NEBULIN 115.
FT	REPEAT	4226	4256	NEBULIN 116.
FT	REPEAT	4264	4294	NEBULIN 117.
FT	REPEAT	4329	4359	NEBULIN 118.
FT	REPEAT	4330	4360	NEBULIN 119.
FT	REPEAT	4365	4395	NEBULIN 120.
FT	REPEAT	4400	4430	NEBULIN 121.
FT	REPEAT	4435	4465	NEBULIN 122.
FT	REPEAT	4471	4501	NEBULIN 123.
FT	REPEAT	4544	4574	NEBULIN 124.
FT	REPEAT	4575	4605	NEBULIN 125.
FT	REPEAT	4610	4640	NEBULIN 126.
FT	REPEAT	4645	4675	NEBULIN 127.
FT	REPEAT	4680	4710	NEBULIN 128.
FT	REPEAT	4716	4746	NEBULIN 129.

Query Match 2.1% Score 415; DB 1; Length 6669;

Best Local Similarity 17.7% Pred. No. 5.4e-11; Matches 695; Conservative 463; Mismatches 1299; Indels 1440; Gaps 180;

QY	265	VNKNVWEAITCDASYKSGYFMOSESNTPLFSNPKCGHK----	QKVPYMLDYVPOYLW	320
DB	66	VERAKYIRKKVDSKFTPIAHOSKQWDLFSNKKYKEKRGQYASTDTPELRR-		124
QY	321	FDEWGEFCKRKRIKAKVAVDSRNDKERLYCSHNGHDCITTTWKIGLILDNKCDT		380
DB	125	-----IKKVDOLSEVKYRW-----DGDVANT-----ICHVDEKADIEH		159
QY	381	KCAVEFWLGNQOEAPFKKOKEKYKEIQSYLSNDN-----KPVNN-----INSEYKOFYEK		432
DB	160	AKRY-----SQVSKYLKQNMEDTKYLLPPDAPELVQAVKKNRMSKKLITDEWEA		213
QY	433	LKETOYATNDT-----FLNLNKGKCYKGGIPGSKDITFTNSADKGIIFYRSEYQV		484
DB	214	DKSLFYPYNDSPELRVAQAKALSDVAYKKG--LAEQAQFPLADP-----		259
QY	485	CPUCGYKCDGKITKTHKSDNDREYNNEDYKPPWGVKPTNITVLYSGNEQGITQKLENEFC		544
DB	260	-PD-----IEFAKKVTN---QVSKOKYK-----		278
QY	545	NSSTNYKDKNNOKWECYKQENINRCKLEQNTTEINNDPKIISFHNFEMLVYLLDRTI		604
DB	279	-----EDYENKTKGW-----SETPCEYANAMNDNDNISTRYQODEF-----		317
QY	605	KANDKLTKTINNTHHCIDECNNRNCICFDRVWVOKKEBMSIKKLFKKKNOOSYYSNI		664
DB	318	--NMKQIYMWQET-----PEKKN-----KKAGVAAS-----		344
QY	665	NNLEFGYFVKVMDKLDKDEKMKELMENIRKK--NEFSMLENNRDYLENAIELLDHLK		722
DB	345	-----KVKYKEDYKNGKADYNVLPASENQ-----LRQK		376
QY	723	ETATICKDNNTNACETSHNATNPCKVPRGQTQPTNKIEIAQYFRSAYEARNGH		782
DB	377	AAADALSDKIKYENETKAKSINCEYK-----FKLDTVLQNF--SSDKYKDSYK		428
QY	783	KIKGAHEGIIYKRGGRKDEKDNLCRIIMKSHNRNLFSGNPGDKGTGDIQTRFVVG		842
DB	429	DIUG-----HYVGSFEDPYHSHCMKVTAQNSKN--YKAVEBDKNGKGFEPGT---ITQ		477
QY	843	EME-VDEPHMRKDHEDYIMPPRRRHICTSNLEHLQTDHPLNGNIYDDLVNNSFLGVL		901

Db 478 EYEAIRKLDCKDHTYKVHPDKT-----FLOYVDSV-----LLOAQVNSLOSLN 526
 Qy 902 SAKTEANKI-----IMYKKNL-----KGEVTDPRHOTTICRAIR 940
 Db 527 KAHHESEKFCGHPPTPTAIOIKVNAVNLSDMLYKODWEKRAKFDIVDAIPPLAAK 586
 Qy 941 YSPADIGDIIIRGDLERNNDVYKLOGHLETVGNIIKSLKGNKYNDAPKYL----- 996
 Db 587 ANKNTSDVYKRYKDKNGKMI-----GVLSIND- PKMHSK 625
 Qy 997 -----KLEMMENRAVW-----EAMCDIKLKSGH----- 1027
 Db 626 VAKNODRLKENYEKTKASNNYCETPKYOLDTOLKNFSEARKDL-VYKDVIGHYGS 684
 Qy 1028 -----OSTOSSCYSDHTPLDDYIPQKLMFMWMAWYKVKKEKDK 1073
 Db 685 MEDPYTHCKKVAANOSSDKSYKAEYEDKGCYFOTI-----TOEYDAIK 730
 Qy 1074 EKCEKCKDNGOGCTKESGTCT-----KC----- 1099
 Db 731 -KLDCKDHTYKVHPDKTFTAVTDSVLLLOALNTKOLSLNRYAKHGEHCKHPIPAD 789
 Qy 1100 -----TEACNEYNDITGKKEW-----NISDKYELHE 1129
 Db 790 APOFIORHNAVNLSDNYKODWEKRAKFDIKVDAIPPLAAKANTKTSVYMKKDYE 849
 Qy 1130 QAO-----MSVNSG--TEASTANHIDRNVIEFSLYOONGSKNSGCT----- 1174
 Db 850 KSKGKMGALSTINDPKMLHSKLTAKNOSDRE-----YKDOEKSTYITAPLDMIQ 901
 Qy 1175 -----SDSANVIGNTYENGALHDGTGNDDCOSONEFCEKSDGKNKFAFRDPDH 1231
 Db 902 VYQAKSKOIASDVDKH-----LH-----SYSY--PPD- 929
 Qy 1232 DGACGSGSKETRYQIKTKKAEKDECKT-VNDILK-----ENDGKROVEDC 1280
 Db 930 -----SINDLAKKAYALOSDVEYKADYNSMMKCGWVFGSLEMKAKRASDI 978
 Qy 1281 HPKNSNGPYDQCGNINIVEDPRVCMPR--ROKLCVHFLANDNET-----K 1326
 Db 979 LNEKTYROHPDYL-KFTSIEDAPITVOSKINQAOISDIYAKKEEITIHVNLPRDLQ 1036
 Qy 1327 KLOSOVNLKEAFIKSAAEFESWYKSKDEGNELDEKELKEGIPAPILSMYTTGCD 1386
 Db 1037 FLOAKVNAV-----ISEMMYKADLKOLSKGYDLRTDAIP--IRAAKAAQ 1082
 Qy 1387 YRDFLEGTDISKGH-----EGSKLKE--QIDSLEKNGDOSKSPGKTROEWTEHS 1435
 Db 1083 ASDVOYKDYERAKGKMGVFOSLDDPKLVHYMNAVAKIOSDREYKDYKSKYNTPHD 1442
 Qy 1436 HEIWEAMLCALVKGAKK--DDEFTENYCYNNVAKFSKSTTLEPAKROFLMWTETWDDY 1494
 Db 1143 -----MNVYAAKKAOVVSNVYK-----HSLH 1166
 Qy 1495 CYTROKYLKADVOEKCSNDOLCDTECNKCEDYVYKMKKKWIP-----ODKYKD 1547
 Db 1167 HYL---VLPDAMDELSKNMMQIOSD-NVYKEDYNNMK-KIGITIPIGSLDVEYKAAQD 1221
 Qy 1548 ERDKRRDRGROIGVAVTDYGTNAIDYLNKRFYASCGDPGASAVORNIQ--LEKOA 1604
 Db 1222 ALNKKY--ROHPDYL-----KFT-SIVDSPVYQAKONTKQVSDILYKAK 1264
 Qy 1605 YVADAKKCGGCTIENDKTYNISKDKCGLYEANTGAIKMNKPNNTNLIKELTED 1664
 Db 1265 GEDV-KH-----KIT--MSPLPOFLAKCNAVYIS----- 1392
 Qy 1665 VLFPSRLRIGCFHALDGNVTDPEYKDENGELKRLMEVAATEGYNIQOYKKEKEKEKIKT 1724
 Db 1293 -----DVCYKR--DMHDLIRKGNVNLGDAIPITAAKASRNILASDYKKEAYEKS-K 1340
 Qy 1725 SDAKYSYEVPCSAMKYSYDLDIILIGDNLDEKOKTEENIKLTFKNGYSVGKSD 1784
 Db 1341 -----GHNHGFSLDDPKLVHYMNAVAKIOSDREYKKNYENTTSYHTEGD 1386
 Qy 1785 STTGNGSTARKFEWENKE-----CWNAMIGYKRGGRDNGSGNSARS----- 1830
 Db 1387 MYTITPAKKAQOVATNVANVYKOPLHHYTLPLDMSLEHRRNVNOIOSDNYVDEYNSFLG 1446
 Qy 1831 -----EDLKCGSVPSDDYPMKKNDEBTAVQFLWFAWGEDFCRKEKELE 1879
 Db 1447 IGMPIGSLVEYKVRKADALNERY-----ROHDPYKFTSVPSMOMMLAQHNTKQS 1501
 Qy 1880 KLVGACNDYTCGDNDKRRKCTDCTOYKKT-----SBMKROYKOLK-YGE 1927
 Db 1502 DL-----NYKVGEKLIKRYTID-PELPOTIOAKVNALNSDHAHYKADWKTTRKGYDL 1554
 Qy 1928 MKDKYSEHPAKDAEDAREYLDOLKIKCNKSGDEYKCMKDVSTQRLDGSQNNPA 1987
 Db 1555 RPDAL---PIV-----AAKSNRIASDCYK-----EAYEAKGQOYF-L 1591
 Qy 1968 SLDDPEKRY-----EGKCNQOYPRG--PRVYRBTPEPRVSLISKATSKKEAKTAPPT 2039
 Db 1592 SLDDPKLVHYMNAVAKIOSDREYKKGYEASTKYHTPLDMSV--TAAKSSEVATNA 1647
 Qy 2040 KOPKVENLJTEMRA---QTRRAAQOTRKRTSTATTTESDVTM-VKALISNKPSR 2094
 Db 1648 NYRQSYHHYTLPLDALNVEHNRNMOIOSDMLYKSDTNNMKKGWVPIESLEVEKAKKA 1707
 Qy 2095 GIEGCKNKTGYOY-----KMGCTVGSKRENN 2123
 Db 1708 GEI--LSKKYROHPEKIKFYAMDTEOALNKNKIMDKRYTEKX-----NKDKT 1758
 Qy 2124 GICPPRRKIKLINIOTY-----NYETEKRNDIKENIKCALET----- 2166
 Db 1759 TIHWPDPTOLLSRVNGITMSDKLYKAGWEKKEKGYDLRPAIATKAARASDIASDY 1818
 Qy 2167 -----QFMKLYIENPAENELONGTIPD--EFK----- 2194
 Db 1819 KYKRAYOAKGKHLGFSLEDDPKLVHMOVAKKOSDREYKKGYSKTSFHTPYDMLSY 1878
 Qy 2195 -----RIMYTY-----GDYKDMFG-- 2210
 Db 1879 VAAKSOEVATNAVYVHTYNNLPDAMSEFLAKNMMQIOSDNYKADADPMKIGWL 1938
 Qy 2211 -----TDISNDK-----KITVYNSVTIILNENKRODKKKDELRK 2248
 Db 1939 PLGSLAEKNNKAMEIISEKKYROHPDLYKYSTLDSMNNVLAQNNAKIN--NEHLYK 1994
 Qy 2249 IFWE--KNKKFIMGW-----YGLTYHLDENEKELND----- 2281
 Db 1995 QAMFADTKVHMDPIQIILANAMALINISDKLYKLSSESKRGYDLRPAIPIRAKA 2054
 Qy 2282 -----NYOVN-----DWTKLPISLEFVYKROFLMPTETAEFECK 2318
 Db 2055 SRDIASDYKRYKYNVEKGMVGFRLSEDDPKLVHSMQ--VAKMOSDR--EYKKNYENT 2109
 Qy 2319 RKE-----OLKLEAGCKEYECNSDGTQCEACAVTYONEIKW-----KTEYR 2366
 Db 2110 KTSYHHPADMLSYTA-AKDAQANITN-----TYVKHLIHXYILLPDAMNIELTR 2157
 Qy 2367 QREKFKDKDKK--KYD-----PSTERDIEKATCAHELYLMKLEKCGNKDCSCMOK 2418
 Db 2158 NMRRIOSDEYKODYEMWKGLGWSFAGSLEVEKAKKAEYAS-----DDKYROH 2207
 Qy 2419 PSS-OLPKYTOOSO-----SSDANDPESIDYVPEEENKCECELKSKGSMITKKTIBEK 2473
 Db 2208 PSNFOEKLKILDSMDVLAQONAHNTNKHLYTL--DMNK-----DKTIHWPPTPOLIOLK 2261
 Qy 2474 IPANCYKAAVYLSKAEENMDITLKEF--IPESIKKEKESNNTNNPCPKPYAPD 2532
 Db 2262 QNOTLYSOKLYKIGWE-----ALKKGYDLPVDAISVOLAKASDIASDYKOGYRQ 2315
 Qy 2533 --KYIGR-----NPCENRENRFYVDE-W-----KC-- 2557
 Db 2316 LGHHVGFRLDDPKLVLSMNAVAKKQSEYKKEDEKKTSSPYDMLGVYLAACKQEL 2375

QY 2558 -----YKNSKFEQKRVCPVPRREHMCLEINLEIKERLKSNDYLIKAVRTPARNECID 2612
 Db 2376 VSDVDYKYN-----YLHOWTCLPDONDVQAKVYELQSENLYKSDL-----EWLRGIGWS 2425
 QY 2613 IIKNFENSENCANPICDMFKYSFADIGDLYRGTDMRIGVLPPEYIKLYK-----FEYI 2669
 Db 2426 PLSGLEKKNKRASEISEKRYR-----OPDRNKTSTIPDAMDIY 2466
 QY 2670 YGRMRKNNKRNKYNVQTFERSAMWDMNR-----KDIWMKTCAPAKEDAKLEFKGR 2720
 Db 2467 LATNKNKNSDRILYRRA-----WDKDTQIHMEDTPYIVAKANLINTSKLYMM----- 2517
 QY 2721 MDGFERITLQDKCGHDPVDVYIPORFRMTE--WSEYCKALMELEKFKKSCDH-- 2777
 Db 2518 --GYEEL-----KRRGY--DLPVDAIPKAKASREIASEYK-----GFRKQLOLHNI 2563
 QY 2778 -----CKTSDCKNDYDKCE-----QCTRCQ--E 2802
 Db 2564 GARNIEDDPKMMWMMHAKIQSDREYKRDKEKKTKFSSPVMIGVYLAVKCTOLVSDVD 2623
 Db 2803 YKNEVLMKSL-----FDIOS--NKYK-----ELYEQPIYTK 2832
 Db 2624 YKNTLHOWTCLPDONDVYIHAQAYDLQSDMLYKSDLOMLKIGMTSGSLEDEKNKATQ 2683
 QY 2833 ISTYDHVQNFQKLTQKTFKSCVSESEFSEYLTHTSK--CLNFK--FNENDGSSNIRTAYAEET 2890
 Db 2684 ILS-DHV--YRQHPDQEFKSLMDSIPVLAKNNAITMNRLEATWMDKDTVHIMPT 2740
 QY 2891 PKGYKAGCSTLPSKNLNDPCPDONKCKELOTFTFCCKNDYDNLMNATVLVNSS 2930
 Db 2741 PE-----VLAKONKRVSEKLYGLE-----ARRKGDMDV----- 2775
 QY 2951 DDKNGVLIPIRRRHLCRTPTAYNRKDEIL-----KKILTSASFQGLG 2999
 Db 2776 -----AIPKAKASADIASSEFKYKGRKQLOHITGARINDPDKMMWMMHAKIQSD 2829
 QY 3000 QKYSSEELCEPMKYSYADYSDIIGTDMADTSLSEKIKIETSEATEENKRTWENN 3059
 Db 2830 REYKRD-----FEKWKTKFS-----SPVDMIGVYLAKKCTOLVSDVD-----YKNY 2870
 QY 3060 RQDIWHMILGKYKATSKYVLDSGWOOLPRDETNQFLRML--EEMAKQCKEKKHSHDS 3117
 Db 2871 LHO-WTCLPDONDVYIHAQAYD-----LQSDNMKYSLOMNRGIGWSIGSLDVE----- 2919
 QY 3118 LKTCRPSNEDNEFASSELKPOGCONDIRYISL-----NLIKNTENTENIK--YKOLKD 3171
 Db 2920 ---KCKRATK--ILSDKIYRQ--PDRFKFTSTVDSLEQVLAKNNALNNKRLYTEAMD 2971
 QY 3172 QSSGNIDNKSEENVQSYKSKDSQCALELNDINEIVTGTNNENNEKREYLAFLGLY 3231
 Db 2972 KDTQIHM-----DPEIMLAKONKIN--YSETLKYL----- 3003
 QY 3232 FVEDETHKNVLDGNIKEEQYAPKALYFTPHVDSFYQAPLESTHRVAYQYDPKNDLK 3291
 Db 3004 -ANDEAKK-----GYDLKSD-- 3018
 QY 3292 SSISVIVSALGLTALHMKRKKFSYDLRLILNIPQEGYGMPTLESKNRYIPYRSGPYK 3351
 Db 3019 ---AIPVIAA-----KASRDVI-----SDY-----KXKDG--YR 3042
 QY 3352 GKTYIYEGTSGDEKYMMDLSSDITSSSESEELDINDIYVPGSPKTYT-----L 3404
 Db 3043 KOLGHHHIGARNIEDDPKMMWMMHAKI--QSDREYK-----DEKWKTKFSSPVM 3092
 QY 3405 IEVVLSESKDIPSDDPSPMDTPRTNRFIDEM-----NELKHDPSOYLPTNEPNNNY 3458
 Db 3093 LGVVLAKCOTLVSDVYKNTL-----HEMTCLPDONDVYI--AQAAYDQLQSDNY 3141
 QY 3459 K-----SADIPMNTPENTLYSDN-----PEEKFTIISHDRDLTYGKEISY 3499
 Db 3142 KSDLOMRLGIGWVPIGMDVYKCKRAAEILSDNITRQPPDKLAKFTSVTD----- 3150

QY 3500 NINMSTNNDIPMNAR 3516
 Db 3191 SLEQVLAKNNAIKNKR 3207
 RESULT 4
 ID S230_PLAFO STANDARD; PRT; 3135 AA.
 AC 008372;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE transmission-blocking target antigen S230 precursor.
 GN S230.
 OS Plasmodium falciparum (isolate NF54), and
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_Taxid=5843, 36329;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-Isolate NF54;
 RX MEDLINE=93241227; PubMed=8479460;
 RA Williamson K.C., Crisculo M.D., Kaslow D.C.;
 RT Cloning and expression of the gene for Plasmodium falciparum
 transmission-blocking target antigen, PfS230.";
 RL Mol. Biochem. Parasitol. 58:355-358(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Isolate 3D7;
 RA Bhatti S., Alano P., Luo C., Hansra S., Alkawa M., Carter R.,
 RL Elliott J.F.;
 CC Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: SURFACE PROTEIN.
 CC -1- DEVELOPMENTAL STAGE: GAMETOCYTE.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.1ab-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 DR EMBL; L08135; AAA29734.1; -;
 DR EMBL; L04162; AAA57559.1; -;
 DR PIR; A48584; A48584.
 KW Membrane; Repeat; Malaria; Antigen; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 3135
 FT
 FT DOMAIN 280 304
 FT DOMAIN 379 410
 FT REPEAT 379 382
 FT REPEAT 383 386
 FT REPEAT 387 390
 FT REPEAT 391 394
 FT REPEAT 395 398
 FT REPEAT 399 402
 FT REPEAT 403 406
 FT REPEAT 407 410
 FT DOMAIN 411 442
 FT
 FT REPEAT 411 418
 FT REPEAT 419 426
 FT REPEAT 427 434
 FT REPEAT 435 442
 SO SEQUENCE 3135 AA; 363213 MW; B5F850C87ED9C77 CRC64;
 Query Match 2.0%; Score 381; DB 1; Length 3135;
 Best Local Similarity 18.1%; Pred. No. 7,7e-10;
 Matches 626; Conservative 431; Mismatches 1153; Indels 1256; Gaps 169;
 QY 524 ITVLXSGNEQDITQKLENPCN--SSTNYKDKNNQKWCYTKDENINCKLEQNTIIND 581

[illegible]

Dd	805	FEMPIYKKEIKNKKENKSLGNLYNNSVYYNKKEMKKNFYNNQYVNH-FPSYDOLNLECS	863
Qy	1570	-----NADYL-----NKRFTASGCDKRGSGASVYORNIOLIEKOAYTDADKHCC	1614
Dd	864	ILKEESNNLSTSYLAVYVSIINELENSLEDFEESHVPVPIKKTLOVAKNV--NNNEHYT	921
Qy	1615	TKFEINDOKRYNISKVR-----CKGLVKEANTAIKWKONGPNNYNNKLELTDLPFSR	1670
Dd	922	CDFFDKDKDYVPSYANGKRLFLICKRHLKEDPTTLKC-VNKTQYRPNL-----IPFK-	973
Qy	1671	RLRLCFHALDGNVTDPEVKDENGILKRIMEV-----AATEGVNGOYRK--	1744
Dd	974	-----TLKRCKEVLTAKLDDLOIQOMSEKFEKFNTOAKNLNYPYLLFP	1017
Qy	1715	-----EKKKEKEKITSDAHKSYEPPOSAAKYSFVYLDRIIGIDNLEB-----EKO	1764
Dd	1018	FNHGKTKELNPNPYKHK-----DVKFEQSSVLSPLSABSLGKLNLFDPDQ	1066
Qy	1763	KTEENLKIFKNKGTSVGKSGDSTTGNGSTARKEFWNNEKCVNMAMICGKRGDRDN	1822
Dd	1067	EVVCLTEKIRYLNLSINELSDNNTPFSVTRQVPPYI-DIKEPFYMFCCNNNG--EGN	1122
Qy	1823	SG-----NSARDEDLKRGCSVSPSDDYPMACKNDEGTAQFLMRAEMGEDPCKRKEKL	1878
Dd	1123	IGIVELLISKQEEKIKGCNFEHSLDY-----	1149
Qy	1879	EKLVCACNDYCGNENKRRKCTPDATQYKKFISEMKPOYKQIKKYGKNDKIYSEH-	1937
Dd	1150	-----FENENSSD-----THECLH-----AVENDIIGF-----NCLETHHPN	1162
Qy	1938	-VANDAEDAREYLDKOLKTKICENKSGDCEYCKKNDYSTORLTDG-----NSQNPASLD	1990
Dd	1183	EVEVEVEDAEIYLOPE--NCFPN--YKGLNSYDITILKMAQYNNINNKRPPELK	1235
Qy	1991	DEP-----KEYVGKNCQVPPGPPV-----RRETPS-----	2017
Dd	1236	IPPYULEDEVLSQCTIKQYKIKVITKNDYLLKKEVOSSESTLDDKIKYCHEHFI	1295
Qy	2018	-PRVSLI-----SKATASKREAK-----TAPTRKP--KKV	2045
Dd	1296	NPRVAKTEFDEVETCNKIKTEFPFYIOLIFCAPKDGIYKNIQMYDYIAPTRVPOPKFE	1355
Qy	2046	EN-----LTTEBRATRRRRAAQOTRK-----RTSTATTT	2075
Dd	1336	NNBEHLKLIIPNSEHLTKELILLYNEEKVDLHFVPLPIYIKDIYEENIYCNDSKTMW	1415
Qy	2076	ESDY-GTWVKALISNKRDSRGIEGC-----NPKTYQOYKWCICIVGSKENENGICM	2137
Dd	1416	KNQJGKVIYHITYSKREQ--YKGSFEDENAHMFVYKNTVKNKCIIDAKRKDLIGVC	1473
Qy	2128	PPRRKKL--C-----INNIOYL-----NY-ETENKRDNDIKEAFI	2159
Dd	1474	PSGTLKLNCKDAIVHTNTLNINGILYKNNLANFTYKHOFNYMEIPALMNDIS--F	1530
Qy	2160	KCAAIETQFLMKIYIINPAENDEL-----NCTIDFERKRIYUYGSD--	2203
Dd	1531	KCIQVYDK--KKKNVAKSPGPRVRLALYKLIKIKDKNTVGT--DQNKYLM--TYMDJH	1564
Qy	2204	-----YKDMFECTDISNDKIIITYNSVTYI-----LNENKKKQDKKDELRKIPWE	2252
Dd	1585	LSHKRNLYLKLTFH--DLGKKRPADPDANPSSITESISINESNESPFPGLGVDAEHL--	1639
Qy	2253	KNKPFINGMKIYGLTYLTDENKEKIRNDYQYNDMFKL-TFSLEEFYKRRQFLMFWEM	2311
Dd	1640	-----ILEG--YDTWESLYEOLLEEYI-----YNDIESLELDIOQYVLOVNL-----	1680
Qy	2312	AEEFCNKRKEOILLLEAGCKEYECGNSNDGQTECACAECVYQNFILKKWTEYEROREKEF	2371
Dd	1661	-----KAPRLMAEAOIHHNRHRCODSKN-----NLIVP-----ESL	1711
Qy	2372	KKDKDKKKYKDYPESTEDIEKATCAHEYMLMKLKEJCGK--DCSCMKPESQOL--P	2424
Dd	1712	KKK-----EELGGPNVHICHTYALKLPDLTLYVCP	1741

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; 019769; AAA82889.1; -
DR EMBL; U30872; AAA82935.1; -
DR EMBL; U25725; AAA86889.1; -
DR Genew; HGNC:1857; CENPF.
DR MIM: 600236; -
DR InterPro; IPR001230; Prenyl_site.
KW Chromosomal protein; Nuclear protein; Centromere; Coiled coil;
KW Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat.
FT DOMAIN 14 197
FT DOMAIN 273 769
FT DOMAIN 823 1338
FT DOMAIN 1642 1746
FT DOMAIN 1862 2987
FT DOMAIN 2207 2568
FT REPEAT 2207 2386
FT REPEAT 2389 2568
FT DOMAIN 3015 3032
FT CONFLICT 16 16
FT CONFLICT 250 250
FT CONFLICT 272 272
FT CONFLICT 611 611
FT CONFLICT 1494 1589
FT CONFLICT 1611 1611
FT CONFLICT 1811 1811
FT CONFLICT 2242 2243
FT CONFLICT 2335 2335
FT CONFLICT 2492 2492
FT CONFLICT 2545 2561
SQ SEQUENCE 3210 AA; 367589 MW; 11083324960E4334 CRC64;
(IN REF. 3).
NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
T -> A (IN REF. 2).
L -> Q (IN REF. 2).
G -> D (IN REF. 2).
MISSING (IN REF. 2).
MISSING (IN REF. 2).
V -> A (IN REF. 2).
V -> L (IN REF. 2).
ER -> DG (IN REF. 3).
L -> Q (IN REF. 3).
D -> N (IN REF. 2).
ELNERYALHNDQEAACK -> SSMREMOPTMTKKPV
(IN REF. 3).

Query Match 1.8%; Score 357; DB 1; Length 3210;
Best Local Similarity 18.7%; Pred. No. 1e-08;
Matches 574; Conservative 481; Mismatches 1120; Indels 898; Gaps 151;

DB 605 KWNDLKTCINNTWTHGCDCECRNCLCFDRWYKKEEEMNSIKLFTKKKNIQSYINI 664
7 EKKELIPI-----RTLOKIOELBOQ---LDKLKKEKQOROFOLDSEAPQOTQKVE--- 56
665 NNLFEGYFFKYWDKLDKDEAKKELMENIKRKKNEFSN-----LENNRDYLE---NAIELL 717
57 NEKTEG-----TNLKRNOIRMEICESELEKTKOKISHLQVKESQVNFQSGQNSGKKQ 110
718 LDHLEKETAICKDN-NTNEACETSHNATNPQVPRG-GTQPTNRIKILAOYFRSAVEE 775
111 IEKLEQELKCKSELEKSOQAASQVSLNCPNPQKLETFPLRP-----SQYSSSKYED 166
776 ARNGLHLKAKAHEGIIYKRRGRKDFKDNLCRIIMIKSHNNLGFSGNCPDQKGGDQIQ 835
167 -----LKRYNKEVEE-----KRRLAEYKALQAKAASQTL----- 197
836 TRFVVGTEMEVDPEHMKRDEHDVIMPRRRICTSNLEHLOTDHP--LNGNIYDVLNN 893
198 -----PQATNNHNDI-----ARHOASSVFSMOQEKTPSHLSNSQRTPIRR 239
894 SFLDVLISAKYEANKIIRMYKKEKNLKGPEVNDPKHQTI--GRIATRSF-----ADI 946
240 DF-----SASY-----FSGELEVTPSRSTQIGKRDANSFSGNSSPHL 279
947 GDITRGDLWERN--GDM-VKLGHELTVEGNHKSLSKGGNDKYNDAPRYLLIRENMW 1003
280 LDOLKAONOELRNKINLELRLQGH-----EKEMKGVN-----KFOELQL 320
1004 EANAATWEM-----KCDIKILKDSGHOSSTSCYSTGSDHTPLDDYIPLQRLRWTE 1056

DB 321 QLEKAV-ELIEKELKYNCRDELVATTQYQASRTYA-----LEQKLKITE 369
1057 MAEWYKQVAKK-----YDKLEKCKECKDN-----GOGCTKESGTGCTKE 1101
370 --DLSQOROMASARCSLQDKIKEKEKEFELSQRQSFQTLDEQIDQMRARLQELQ 427
1102 ACNEYV-----DIIGKMEKNIIIDRYKELHEQAKMSVSGSLASSTAKNHIDRVI 1155
428 ARNMHVLADELDKLTSYKQOLENNLEBKQ-----KICRAQAVQASQIKENELRRME 482
1156 EFLSELVQONGSKSNKSGTSDSAVIGTNTYENYGAYLHDGNPDQOSQNEFCDEKSD 1215
483 EMKKE--NNLLKSHSEKARE--VCHLEALKNIKQCLN-----QSON-FAEBKA 528
1216 GDNKRYAFRDKRQHDQACGCKSSKSPRQVITKKAAE--EKDTCKTVNDLIKEND 1272
528 KNTSOTMLRDQF-----KIQQENSITLEKIKLAVDLEKORDCS--ODILKKRE 578
1273 GKQVEDCHPKKNSNGYDWCQGNINLVEDPRVCMPRRQKLCVHFLANDNEIKLOSQV 1332
579 --HHLEQLNDKLSK-----TEKESRALSL 602
1333 NLKEAFIKSAAEITFSWYKSKDGEENELDKELKEGKIIPPAFLRSMFYTGDRDPLF 1392
603 ELKR-----KEYE-----ELKEEK----- 616
1393 GTDISKGEGSKLKEQDISLFFKNGDOKSPNGKTRQEWMTSHETWEAMCALVKGAK 1452
617 -TLPSCMSENEKLTLOESENKQSLNLETLCKQOKISHMYNRVLTLEMD--R 672
1453 KDDFTENTGYNNVYKFSKSTLLEEFKRPQFLRWLTWEYDYCYTRQYKLDVQEKCSN 1512
673 ENLSVEIRNLNNV-LDSKSVETQKLAVMELQKAEBSD-----QKHQKEINMCLKT 725
1513 DQKCDPTCKNKCEDYV-KYKKKKKEMIPQDKTYKD----- 1547
726 SOL-----TSQVDELEHKQQLLSNEIMDKRCYODJHAYESIDRLDKSKDASIVTNE 779
1548 -ERDKRRFRDQ-----HICVMTDYGTNATDYLRKRKFTASGDKPGSASVQRIQILE 1601
780 HQRSILAFDQPRAMHSHANITIGQSWPSSRSCLBAD--QSPKSAIILQNRVDSLEF 837
1602 -----KQATYDADKHC-----GCTFIENDDKTYN 1626
838 SLESOKQWNSLQKQCEELVQIKGEIEENILKAKBOMHQSFVAETSQRSIKLOEDTSHAQ 897
1627 I-----SKDKCKGLYVE-ANTGAIKQNGKPNV--NNLKE--LEEDVLFPSSRL 1672
898 VVAETLSALEKKEKELQDLNDKYETBOAEIOELKSHLLEDLSIKELQILSETLSLEKE 957
1673 RICHALDGNVTPDEVNDENGLRRLMEVATGEGYNIQO-----YKKEKEREKIKTSD 1726
958 MSSIIISLNKREIE-ELQENGTLKEINASLNQEMNLIQSESPANTIDREKSIISLSD 1016
1727 AHKSYEV--PPGAMKYSYDLRD-----IILGIDNLEDKQTEENLKR 1770
1017 QYKQEKILILQRCBETGNAVEDLSQYKAAQEKNSKLECLINCTSCENRKKMELQJKE 1076
1771 IFNKGTSVKGSDSTGTGNSTARKFFMNENK-----ECVNNAMICGKRRDQNSG 1824
1077 AFAKEHQEF-----LTKLAFERRNONMLELTVQAL-----KSEMTDNQ 1119
1825 NSARSDBLKCGSVPSDDYPMGKNRDEGTAYQFLRWFAEMGEDCK--HKKELEKLY 1882
1120 NSKSEAGLQELQELTLEKQNKQKQEVND-----LLQENQOLKVMKTHQONLE 1170
1883 GACNDYTCGNENKRRKCTDQCYKFKFISEMKPQYEQIKKAGKKNKDIYSHPAKA 1942
1171 SEPIRNSVAKERESERQOC-----NFKRQMDLEKVE-----ISLDS 1205
1943 EDAREYLDQKLKICEK-----SGDCEYKCK-----DVSQRLDQNSQNNPABL 1989

Db 1206 YNQLV---QLEAMLNKELKLOESKEKELCLOHETIRGDELTSNLDMSO----- 1256
 QY 1990 DDEPKVEKNCQVP-----RGPRV--RRETPSPVSLISATASK-KEATAPPTQ 2041
 Db 1257 -----BISGLKDEIDAEKRYISGPHELSTQDNNAHLQCSLOQTJMKLNELEKICEILO 1311
 QY 2042 PKREVLUTTEM---RAO--TRTRAAOQTRKRSTATTESVDYGVKALKSLKPPSRG 2096
 Db 1312 AEHYE-LVLELNDRSREJCITATRKMAEVEGKLLNEVKILNDGSLHGLVEDIP---GG 1367
 QY 2097 IEGCNPKTYGOYKMGCIYKSKENENGICMPRRKLLCINNIO--YLANETENKRONDI 2154
 Db 1368 EREBOJNE--QHPVSLAPDESNSYEH-LTSLDKVEQMHAELOEKFLSLQSEKILND- 1423
 QY 2155 KEAFIKCALETQFLMKYIIPENPAEN-----ELON 2186
 Db 1424 -----OHCSMSKSELOQVVDLSKAEMLVLTNLNFOGDLVKEQMLGLEGLVPSLS 1478
 QY 2187 GTPD-----EKRIMTYTGDYKDMFGTDISNDKILITYN-SVTITLBNKK 2235
 Db 1479 SCVPDSSSLSDSSFYRALBETQDM-----SLSLNEGAVANSQCSVDEFGCSLO 1532
 QY 2236 KRODKKDE-----ELKLFMEKNKF--IWEGMVGLYH-LDNEKEKIRONYOY 2285
 Db 1533 TYVDSLKAENLVSTLRNPOGDLVKEMLGLEGLVPSLSSCVPDSSSLSDSSFY 1592
 QY 2286 -----NDWTKLPLEEFVRRPOFLRMTEWAEFCNKRKBDQLKLENGCKEYDNGS 2338
 Db 1593 RALLEOTGDMSLTS-NEGAVANSQ--CSYDEVCSLSQEBENLTRK-----ETPSA 1640
 QY 2339 NDKOTBCAECYTONFIKKMKTEYEROKERKKDKDKKYDYSTENDIEKACAN- 2397
 Db 1641 PAKGVELESLECYVYOSLEKLEKEMSO--GIMKKEIQLEOLLSER--OELDCLRK 1696
 QY 2398 EYL--NMKLELCGNKDCSCMQRPSSQLPKTQ-----QSOSDANDMP---ES 2441
 Db 1697 QYLSBNQWQOKLSTVLEMSKLAEMKOTEDLSLEVARIQLOGLDSSSLGIDT 1756
 QY 2442 LDYVPEFNKCEPELSKSKSMITHKITEPKIPMNCVEKAAVYLSKEAEN--NMDITLK 2499
 Db 1757 EDALQGRNESC--DISKE--HTSETTE-RPKKHVHQC--DKDAQDNLNDIEKI 1805
 QY 2500 EKTIPIESTFEKESKNS-WNNNPPCDPK-----PVA--PDKYIGRANPC 2541
 Db 1806 TETGAVKPTEGSGEQSPDTNYPEDEKQSGSECSISELFSGPALVMDLGNEDT 1865
 QY 2542 ENR-----EENRFKVD--YEMKCYKSKFYOE-----KRCVCPPR 2575
 Db 1866 HNLQVKETSNENLKLHYIEDRKRVESSLNEMKELSKLHLOEYQMTKIEACT--- 1922
 QY 2576 REHMCILNIDEIKIERLKSNYL-----LKVARTARNEGID-----IKNFN 2618
 Db 1923 -----ELEKIYGLKESNDSLEKLEYPSCHOELOVETSEGLNDSLEMHADKSSR 1975
 QY 2619 SENGCANPLCDPMKYSFADLGIVGCTDMLRIGGILPPEIKLYKVFETTYGKRNKK 2678
 Db 1976 EDGDVNAKANDWKERFLDVEN-----ELSRIRSEKASJE-----HEALTYE- 2018
 QY 2679 GRNKYNDVOTFRSAMDANRDIKAMATKAPK-----DAKLFKGMGDFEERTLLIO 2733
 Db 2019 --ADLEVYOT-EKLCLEKDNENKOKYIVCLEEBELSVYTSRNOLRGDLMSKKTTRALDQ 2075
 QY 2734 CGHKDDPVDYDIPORFMTWTESEYCYKALMELEKFKKSCDH------KTSDR 2783
 Db 2076 LSEK-----MKRQTOELSHQSECHIQVAVAEVKEKTELL 2112
 QY 2784 CRNDYDNKCEOCKTQCEYKKNVLMKSLFDI-----QSNKYKLY---EQPIY 2830
 Db 2113 QTLSSDVSELKDKTHLOEQLSLEKDSQALSTKCELENOIALNKEKLLVKESSSLQ 2172
 QY 2831 TKISTDYHONFOKLTFTKS-ECSVESFESEYLETSHKCLNKFENEDGSSNRTYAFEE 2889
 Db 2173 ARISESDY-----EKLNVSKALRALVEKGEFALRLSSIOEBVHOLRGIEKLRVREAD 2227

QY 2890 TPESYBACSTLPKSNPLDNCPTDQNKDCGKELQFTFCSKNDYNNIMNNAVYLNS 2949
 Db 2228 EKKOLHIA-----EKLKEREND-----SLDKRYENLER-----ELQM 2261
 QY 2950 SDNKGVLIPRRRHLCITRITAVNRRKGKELKKLL--LTSAPGOLGOKYKSEEE 3007
 Db 2262 SEENOELVI-----LDAEN--SKAEVETLKQIEEMASLVFELDLVTLSEKE 2309
 QY 3008 LCFEAKYSAVADSIKIGTDMADTSLSEKIK--KIFETSNATENRKTWENNRQIY 3064
 Db 2310 NLTKQIOENOGQSELDKLSSFKSLBEKEQAEIQKESKRAVE--MLONOLKELN 2365
 QY 3065 HAN--LCGYK--IATSKVTLDEGWCQLPKDEETNQF-----LRWLIE--MAKQAC-- 3108
 Db 2366 EAVAAALCGDQIEIMKATEOSID-----PRIEEOQLRNSIEKLRARLEADERKQOLCVLOQ 2419
 QY 3109 -REKRVSDSLK-----TKPRSNEN--FEAS-----ELLRO--PGQNDIRKY 3148
 Db 2420 LKSEHRAIDLKGRVENLERELIARTNOHAALAEANSKGEVETLAKIETGQSLRG- 2478
 QY 3149 ISLNL-IKNTMENL-----NIKYOLOKQSSGNIDNKPSEENVOSIKXSQCA 3198
 Db 2479 LEIDVYTISEKDLNLELOKEQERISELEITINSFENILOEKQEKYQMKESSTAMEX 2538
 QY 3199 L--ELNDINEIVTGKNN-----ENNEKEY--LKK--LYPGLYVEDETHRNV- 3242
 Db 2539 LOTOLKEINERVAALHNDQACRAKEQNLSSQVECLELEKQALQGL-DEAKNVIY 2594
 QY 3243 -----LDGNIRKEE 3251
 Db 2595 LQSSVGLIOEVE 2607

RESULT 6
 PVDR PLAYS
 ID PVDR PLAYS STANDARD; PRT; 1070 AA.
 AC P22290;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Duffy receptor precursor (Erythrocyte binding protein).
 GN PVDR.
 OS Plasmodium vivax (strain Salvador I).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxId=126793;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91187056; PubMed=1849231;
 RA Fang X., Kaslow D.C., Adams J.H., Miller L.H.;
 RT "Cloning of the Plasmodium vivax Duffy receptor.";
 RL Mol. Biochem. Parasitol. 44:125-132(1991).
 CC 1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
 CC DETERMINANT.
 CC 1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC 1- SIMILARITY: HIGH, TO P.KNOWLEDGE DUFFY RECEPTORS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M61095; AAA63423.1;
 KW Malaria; Receptor; Glycoprotein; Signal; Transmembrane.
 FT SIGNAL 1 20
 FT CHAIN 21 1070
 FT DOMAIN 21 1007 DUFFY RECEPTOR.
 FT TRANSMEM 1008 1025 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 1026 1070 POTENTIAL.
 FT CARBOHYD 183 183 CYTOPLASMIC (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).

[illegible]

Query Match 1.7%; Score 337.5; DB 1; Length 1153;
 Best Local Similarity 18.8%; Pred. No. 2.3e-08;
 Matches 237; Conservative 168; Mismatches 434; Indels 425; Gaps 53;

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 626 626 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 722 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 847 847 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 856 856 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 900 900 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 910 910 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 935 935 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1153 AA; 130471 MW; 6497BD16CE7BEC01 CRC64;

QY 2382 DYSTERDIEKATCAHELYNMLKELCGNKKDCQMKPSSQLEPKTQOQSSDANMPES 2441
 DB 40 EYKKNNGYKLAH-GHHYMD-----NDQIERMLQGTDRSRRAKTEEN 80
 2442 LDYVPEEFNCEPELSEKSGSMITHRTKITEPKIPMNCVERAAAYLSKEAEN----- 2493
 DB 81 VKY-----KYNVEELNTEKTEQTKGRIN-----RIKESYEAQNVADNNTIDDKANG 128
 QY 2494 -----MDITLKEKPIP-----TESTKE--KESKMSWTNNNCPDKPKPY 2529
 DB 129 EYKTDNKTNGEGARMMVMALDYDISGHPDGIIDNVVLEGTEDGEGFNLSKSGDHPY 188
 QY 2530 APD-----KYGRNRCENREERFRVDEEMKCYKNSRYEOKRKYCV 2573
 DB 189 RMNRKERMCSGVINOTFLQKNVWR--CNKKRGRTR--DWDC-----PRKQVCIP 236
 QY 2574 PRREHMCLEMLDEI-----KIERLDSNYLLKMYR--TARNEGIDIILKNF--SENG 2622
 DB 237 DRRAYQLCMKELTMYNNVNTKTHSHNDITFLKLNKEKLYDAEAGDILLKKNVYSED- 295
 QY 2623 CANNPICDITMKYFADLGDIVRGTDMIRIGYLPVEIKLYKFEYIYGKMRKKNKGRK 2682
 DB 296 -----LCKDKWLSLEDFGIMGTDEGI-GYSQVVENMLRTYF-----GT 335
 QY 2683 YNDVOTERSAMPMANDRDKAMTKATCAPEBDAKLEFRKGRMDGFERITLLDQCKGHKDDPV 2742
 DB 336 GRTYQLDKRRKMMWESKRYIWEATY-----LSVKKKLNGYSAM--NC--KEDVQI 380
 QY 2743 DDIYPOFRMTWMESEYCCALMELEKEFKSCDH-----CKTSRCKNDYDENKC 2793
 DB 381 -NVEPQIYRIRIRMGMDYSELPEQQRKIKKCDRLIYTNLILCIMS----- 427
 2794 EOCKTQCOEYKKNVLMKSLFLDIQSNKYELYEQPIYTKISTYDHOVQYQKLTREKSEC 2853
 DB 428 -PCNDSCKLVDOWITRRKKQMDVLSYRFSVKKGOIT-----ETENITTAVIDILKOE- 478
 QY 2854 SVESFSYLETJSCNLNFKENENDSSNIRTYAFETPKSYKACGCTL--PSKNPLDNC 2911
 DB 479 -LNGFENWME-----NEINKKRDNV-----YIDICACADEPKMN-- 512
 QY 2912 PTDOKNDCGKELOTFTFGSKNDYNDMLDMNNAVLYVNSDDNKGVILIPRRRHLCRTPIT 2971
 DB 513 -----TQEHKKTKS-----APKLETRQSHSTIQPMS 539
 QY 2972 AAYNRK--GDKELKKKLLTSAFSOGOLGQKRSBELCFEAMK-----SYADYSD 3022
 DB 540 SSGAEVVOGD-----LAHGNINDAAKSTTD--EAKGDDGONGNOTVASEN- 583
 QY 3023 IIRGTMDMDTSLSEKIKIIPETSEATENRKTWMENNRQIMHAMLCGYKATSKYVLDE 3082
 DB 584 -IKGTDNIE-----NEAAKNVDT-----YKFTYERSADTR 612
 QY 3083 GWCQLPRDEETNOFLMLLEMAQACKKKHVSDESKTKCPRSNEDNFEASELLRPGCQ 3142
 DB 613 GADITETGE-----EKLNITSYSGSSE--ITYKENIPGSGIV 647
 QY 3143 NDIRKATISLNLILKNMENLNIYK--QLAKDS-----SGNIDNKPSEEN-VQSYTKSK 3193

DB 648 KDV-----SAAVENSNPETHKRTIFEDSKDONSSENSGSEKFKATSNPTTEAVESS 701
 QY 3194 DSQALENDINEIYVGTGKNNENNEFEKLYKRTYPGILYFEDETHKNHYLDG----- 3245
 DB 702 SAGQVOEBSAHSVMTGDNSTISAASTDGSSGDKRVESLTSIEMADGGDPYQGSLL 761
 QY 3246 -----NIKEEQTVRPKALYFETPHVDSFYQALPSTHRVAQYDPKN-D 3288
 DB 762 WNLNDPSVAGAGGKSHIKTEE-----NEGSAEIDDKNDV 797
 QY 3289 ILKSSISIVY-----YSALGLIALHFMKKKFKSSVDLRIILIPQGEFG 3332
 DB 798 IAOQRTATITTEVOPEBPDLSDTDNGVNPBSG-----NKQNGCATALSGAESLESNEV 850
 QY 3333 MPLESKNRIIPRSGPYRG--KTYIYMEGDTSGDE--DKYMDSDSDITSESEYE 3387
 DB 851 HKTIDTHTGLEKNKNGENEDFOKHDPNMNDMLNDOTSSDQTSDDQTSNOTSSDQTSN 910
 QY 3388 LDINDIYVPSPKYKTLIEVLEPSKRDIPSDPTPSNDPRTNRFIDDENNELKHDFVSG 3447
 DB 911 QTSDD-----QTSDDQSSDQTSDDQTSDDQTSDDQTSDDQTSDDQTSDDQTSDD 953
 QY 3448 YLPNTEPNNNYKASADIPMTEPTIYSDN--PEKFPYIISHRDLYTKE--ISTYN 3502
 DB 954 RNEPISSEDMSSKGDPMRNSNELYSHNMLNNRKLINDOYERHDVAKATREKILIMSEVN 1013
 QY 3503 MSTN 3506
 DB 1014 KCNN 1017

RESULT 8
 MLP1_YEAST
 ID MLP1_YEAST STANDARD; PRT: 1875 AA.
 AC 002455;
 DT 01-OCT-1993 (rel. 27, Created)
 DT 01-JUN-1994 (rel. 29, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Myosin-like protein MLP1.
 GN MLP1 OR YKR095W OR YKR415.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_Taxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C;
 RX MEDLINE=93247549; PubMed=8483450;
 RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
 RA "A new yeast gene with a myosin-like heptad repeat structure.";
 RL Mol. Gen. Genet. 237:359-369(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94205265; PubMed=8154186;
 RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
 RA Remacha M., Jimenez A., del Rey F., Ballesla J.P.G., Revuelta J.L.;
 RA "The complete sequence of a 15,820 bp segment of Saccharomyces
 RT cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
 RT new open reading frames.";
 RL Yeast 9:1349-1354(1993).
 CC -1- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
 CC REPAIR.
 CC -1- SIMILARITY: SOME, TO THE TPR ONCOGENE.
 CC -1- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 DR EMBL: L01992; AAA34783.1; -
 DR EMBL: X73541; CA51948.1; -
 DR EMBL: Z28320; CA82174.1; -
 DR PIR: S38173; S38173.
 DR SGD: S0001803; M1P1.
 KM Coiled coil; DNA repair.
 FT DOMAIN 69 487 COILED COIL (POTENTIAL).
 FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
 FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
 FT CONFLICT 301 301 R -> A (IN REF. 1).
 SQ SEQUENCE 1875 AA; 218455 MW; 683AD034C9066867 CRC64;

Query Match 1.7%; Score 332.5; DB 1; Length 1875;
 Best Local Similarity 19.1%; Pred. No. 7.1e-08;
 Matches 445; Conservative 350; Mismatches 897; Indels 637; Gaps 100;

QY 1321 NDEIKKLGQVNLKEAFISAAAEFF--SWYYKSKG-----EGNEIDKL 1367
 Db 4 HDPMESLQNGENSE---RLNALSFQGLSQVKSFGDGVVYKHLNDKLLQFNELKSE- 59
 1368 KEKIPPAFLRSMFYFGDYRDFLFGTDISKGHGEGSKLEQIDSLFKNGDQKSPNGKTR 1427
 Db 60 -NKVTVYVF-----DELKASSLKIDG---LKTENENYIRENDKIR---KER 99
 QY 1428 QEWMTSHSHIEMAMICA--LVATGAKKDDFTNYGYNVKSDDKSTTLEFAKRPQFLR 1485
 Db 100 NDFVVFESVENKMKLISELEFVKRRKLDLVEKKETOSNOORLIKIDLERLKELEVR 159
 QY 1486 WLTEWYDYCYTRQKYLKDVOEKC--SNDOLKCDTECNKCEDVYKMKKKKEIPQ- 1541
 Db 160 VENNRSNSECCKLRSTIMOLETKQOQYITND-LNSRTELEKTKOE--LTLLSNNMLEKE 217
 QY 1542 ----DKY--YKDERDKKRPD--RQHTGVWVTDY---TGTNADYLNRRKFTASCGDK-- 1586
 Db 218 LRSKNBOYLSTROKTDKVLDIRNELNRLNDFOMERTNNDVLYQNNNELSKSLQEKLE 277
 QY 1587 ----PGASVVOYRNLTLEKQAYYDADAHGCGCTFIENDKTYTINSSKD 1631
 Db 278 IGLGSLNSEKQEPFAESKLRQVLDLESQ-----LNAVK 314
 QY 1632 KCKGLVKEANTAI--KWQNGPNYNNLKE--LTEDVLFPSRR--LRICFHALDGYNT 1684
 Db 315 EELNSTRLEINTAKVLADSKKQTPENEDLKEQLTKEKLAQCEKECRL-----SST 368
 QY 1685 DPEVDENGELRKLMEVATEGYNL-GQYKKEKKEKIKIT--SDAKYSTEVPQCSAM 1740
 Db 369 DDAEDENELSKA---SSDPIFLKQLIKERRKEHLQNOIEFTFYELEKRYPIINSF 424
 1741 KYSF---YDLRDLILGIDNLEDEKQKTEENLKIFNNKGTSVGKSGDSTGNPESTARK 1796
 Db 425 KERTDMLLENELNNAALLLEHTSNEK---NAYKELNNAKNOKLYECENDLQ---LTKOR 477
 QY 1797 PFNNKNECVNNAAMICYKRGHGDGNSGNSASDEDLKKGSPVDDDYPMGKNDEGTA 1856
 Db 478 L-----DLCRQIQLYLLITNSVSNDSKGPLKEE----- 505
 QY 1857 YQFLRFAMGDEDFCKHKEKELEKLVGACNDYTCGDNDKRRKCTDACTQYKKPISKEKP 1916
 Db 506 IQFIONIMQ--ED-----DSTTES--DSQKVVYTERLVERKKNIT----- 540
 QY 1917 QYEQIKKRYGENKDKIYSEHPVANDAEADAREYLDKQLKRICENKSGDCEYVCMQ-VSTQ 1975
 Db 541 ----OLQG-----KNAE-----LLKAVRNLAADKLESEKKSQSLQ 572
 QY 1976 RL---TDSNGSNMPASLDDEKPEKVGKCNCOVPRGPPRVRRRETPSPRVSLISKATASKE 2032
 Db 573 KIESETVEAEAITLKSEKMLE-----SRIEELQKELEE 609
 QY 2033 AKTAPPTKOPKVENLTETEMPAOTRTTRAAO---OTRKRTSTATTESEVGVYKAILLS 2088
 Db 610 LKTSVP--NEOASYSNVITIKQLETKRDLSEVQVLDQTRISQITRESTE----- 656

QY 2089 NKPDNRGIGECNPKTYGQYPMWGCIVGSKENENGICMPRRKRLCINNIOYLNET-- 2146
 Db 657 -----NMSLNKEIOD 667
 QY 2147 --ENKRDNDIKKAFICAAI--ETQFLMKYTIENPAENELQNGTIDPEFRIRIYYYG 2202
 Db 668 LVDSKSDISIKLGKESKSSHILAEERKLLSNLIDLTKAEN-----DQLKRPDYL-- 717
 QY 2203 DYKDFEFGDISNKKIITVNSVTTLNENKRRKQDKKDELRKIFWENKKEFWGGM 2262
 Db 718 -----QNTILKQDSKTHETLNEVYSCSKSISIVETE 748
 QY 2263 IYGLATYHLTENEKEKIRNVOYN---DNTKLTLP--SLEEVKRPQLRW-----ETEMA 2312
 Db 749 LNLN-----KEQKRLVHLEKLNKQELNKLSPKSDILRLWVQLQTLQKREBDELEET 801
 QY 2313 EEFCKRKREQL-----LKLKAGCKEY-----ECGNSDNGTQCAACVATYQNFIKW 2360
 Db 802 RRSCKQKIDLELDALSELKETSQGDHHLKQLEEDNNSN-----IETYNKIEAL 851
 QY 2361 KTEYE---RQREKFKKQDKKRYKDYPTERDIEKATCAHEYLNNKLIKELGKNDQSCMQ 2417
 Db 852 KADYESVITSVOSKQTDIEKLOYK--VKSLEKEIEBEDKIRLHTYVNMDETI--NDSLKE 908
 QY 2418 KPSQLPKTTOOSDANDPESLDYVPEEPNCECEPLSKKGSMTHTKITEPKIPAN 2477
 Db 909 LEKSKINLTDAVSQKEKEDLYETTSOSLQOTNS-KLDE-SFKOPTNOIKMLTEKTSLE 966
 QY 2478 CVERAAVYLSKAEANNMDITLKEKFIPIESTREKESKSNMTNPNCPKRYADPKYIGR 2537
 Db 967 --DKISILKEQMFNLNNELDQKGMEREKADFKRISILQNN----- 1007
 QY 2538 RNPCEENRENFYDYDEKCKY--NSKFYQKRAVCYPPREHKLNLNDEIK-TERIKD 2594
 Db 1008 -----NKEVEVKSSTESKSLIONDDQOTIYANTOANNTEOBLQAHAVSKTISELR 1062
 QY 2595 SNYLLKAVRTARNEGIDILNFSNEN---GCAMPICDTMKSFADLDIYVGTMLR 2650
 Db 1063 QLHTYKGGVQKTL-NLSRQLENALKENKMSWSQKESLQDLQDSNRIDLSQNKLLY 1121
 QY 2651 IGGYLPVEITLYVFEYIYCKMKRNNKGRKRYNDVOTFFSAMDANRNDIWMATCKAP 2710
 Db 1122 ----DQIDITYAAD--KEVNNSTNGPLNNILTLR-----REERDILTKYVAE 1165
 QY 2711 EDALFRGRMDGFERITL-----IQDKGCHKDDPPVD-----DYIIPQFRMT- 2754
 Db 1166 RDAKMLR-----QKSLMDVDELQDARTKLDNSRVKEMHSSIQOHDIKELNQLNL 1218
 QY 2755 -----EMSEYCYALMELEKFKKSCDHCTSDRCNDYDENKCEQKTRCOE 2802
 Db 1219 LRESNITLRLNELENNKKKELESDLDLKNONVAPISEITLTKYSQOEKEQELAKKE 1278
 QY 2803 YKNFVLKWSLFDIOSKRYELVEQPIYTKISTYDVHQNVOVKLTFKSCSVSESEYL 2862
 Db 1279 ----VHRMKK-----RSODILEKHQ-----LSSDVE-----KSELEIENKEEL 1315
 QY 2863 HETSKC---LNYKFENEDGSSNIRTYAFEEETPKSYKACGCTLPKSNPNDPCPTDQNKOG 2919
 Db 1316 ENKREQAEAEKFPNRLRQAGRLKTSKLSQSLTEQVNSLDAKNVYLNENLSISEMA-R 1374
 QY 2920 CKELQTFPFSKDYDNNLDNNNAVYLVNLSDDNKGVLLPPIRRRHLCSTRPITAYNRKGD 2979
 Db 1375 ITELQNAKVAQGN--NOLE--AIRKLEDAEKASRELOAKLEESTSYESTINGLANE 1428
 QY 2980 KEILKRLTLTAPSGOLLGQKYSSEEL--CFEAKYSYADYSIDTIKTDMMQDLSLSEK 3037
 Db 1429 ITLKEIEIKORIOQLOLQATSANEQNDLSNYESKAKSFEE--DKIKIKKEKTOEVNEK 1486
 QY 3038 I---KKIETSNEATENKRYTWENNRRQIWMALCYGIATISKYVLIDGCMQLPDEET 3093
 Db 1487 ILBAOERLQPSNINNEIEIKKWESB-----HGOVSQAIRABEALTKR-IRLPTLEKI 1540
 QY 3094 NQFLRWLIEMAQACEKKEKHYSDSLTKCPRSNEDNEFEASELLRQPGCONDIRKYSINI 3153

DB 1541 NKTIERKEKEFEERK-VEERIKS-----MEOSG-----EIDV 1574
 QY 3154 LKNTME-NINIKYKOLKDOSSGNIDNKPSEENVOSYIKSKDSQALNELNDINEIYVGT 3212
 DB 1575 VLKQLEAVKQOKLEENYNNKKLOELKDVHSHSI-SDDRDKLRA-----EISRLR 1629
 QY 3213 NNENNEPEVKLLYKLYFVEDETHKNHVLDCNI-----KEEQTVRKALYFTTPHV 3266
 DB 1630 EEFNNELQAIKKRSF-----DEGQQAAMKTTLLERLAKMESQJSETKQASPEPS 1682
 QY 3267 DSFYQAPLSTHVAQD---PKNDILKSSISVIVISALGLLHFMKFKKFSVYDLRI 3323
 DB 1683 VNNVQNPVLGLPKRIEENSNPFPNLSGE-----KLKLNKSS----- 1722
 QY 3324 LNPQEGYCPMTESEKRYI-----PYRSGPYKGYIYVEGDTSCGDEKMDLSSDI 3378
 DB 1723 ---SGFNFPTSPSRKHQNDNDKRESJANTDPPTHLEPSPNIPASGL--ISSST 1776
 QY 3379 TSESEYEELDINDIYVPGSPKYKYLEVLEPSKDI-----SDTPSNDTPPTNR 3431
 DB 1777 LSTDTNDEELTSSNNPAOKOSSNNNOSEEDTEKKKEGEPYKRGALIEQTKSNRPP--- 1832
 QY 3432 FIDDENNELKHDFVSOYLPTPEBNNNNKADIMNTEPNTLYSDNPEK 3480
 DB 1833 ---IDEVGLKND-----EDDTENIN-----ESKRKTEDEBEK 1864

RESULT 9
 PVDG_PLAKN STANDARD; PRT; 1070 AA.
 ID PVDG_PLAKN STANDARD; PRT; 1070 AA.
 AC P50494;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Duffy receptor, gamma form precursor (Erythrocyte binding protein).
 OS Plasmodium knowlesi.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NC NCBI_TaxID=5850;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92357776; PubMed=1496004;
 RA Adams J.H., Sim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.;
 RT "A family of erythrocyte binding proteins of malaria parasites";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).
 CC -1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP DETERMINANT.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: HIGH, TO P.VIVAX DUFFY RECEPTOR.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC EMBL: M9695; AAA29604.1;
 CC Malaria; Receptor; Glycoprotein; Signal; Transmembrane;
 KW Multigene family.
 FT SIGNAL 1 21
 FT CHAIN 22 1070
 FT DOMAIN 22 1003
 FT TRANSMEM 1004 1025
 FT DOMAIN 1026 1070
 FT CARBOHYD 134 134
 FT CARBOHYD 134 134
 FT CARBOHYD 179 179
 FT CARBOHYD 676 743
 FT CARBOHYD 743 743
 FT CARBOHYD 785 785
 FT CARBOHYD 936 936
 FT SITE 279 281
 CELL ATTACHMENT SITE (POTENTIAL).
 POTENTIAL.
 DUFFY RECEPTOR, GAMMA FORM.
 EXTRACELLULAR (POTENTIAL).
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CELL ATTACHMENT SITE (POTENTIAL).
 POTENTIAL.

SO SEQUENCE 1070 AA; 120931 MM; 703D68811BC11B50 CRC64;
 Query Match 1.7%; Score 327; DB 1: Length 1070;
 Best Local Similarity 18.1%; Pred. No. 6,3e-08;
 Matches 240; Conservative 195; Mismatches 467; Indels 424; Gaps 60;
 QY 1014 MKCDIKLKSGHOSQSSYSGSDHTPLDDYIPQKLRMTAEAYCYVQKKEVDKL 1073
 DB 35 LECENEYVKNENYKLTATGHH--YMDNDQIE-----RWL-----QGT 69
 QY 1074 ECKEKCXNDKNGGCGTRESGTGCTEACNEYNIDILKKEQONI--ISDKYKELHEQA 1131
 DB 70 DRSRRKATIEEN-----YKTYVNELTWKTEQMKRGK- 100
 QY 1132 QMSVNSNGIEASSTAKNHNDRNIEFLSELYQNGKSNK-----SGTSD 1177
 DB 101 --RINRLKSTYEAQVAVANNYIDKANGKEYKTNDKTNKGEGARNNVMDLYDISGSGP 158
 QY 1178 SVAIGTYYTENYAVL--HTGNFDDCQSONEFCDEKSDKDNKFAFRKPDHDGACG 1236
 DB 159 DGII-----DNVELLTDEGNF-----LKNSSKGDHPYMKRKRKSSGA-- 200
 QY 1237 CKSGSKPTVQIKTKKAEKQTECTVNDILKENDGKKQVEDCHPKNSNGYPMQCGN 1296
 DB 201 -----INQIFLNNYMDKND-----KRRGERNDHDCPT 229
 QY 1297 INLYVEDPVCMPRRORLCVHFLAN-----DNEIKRLQSQVNL-----KEAFI 1339
 DB 230 ---EKDVCIPDRRYQLCWMEITNLVDTDFHSDIIFPKYSRRRLIYDVGGRGDLL 284
 QY 1340 KSAAEFTFFSWYYIKSKDGEENELDKELGKGLPAPLBSMFTYFGYRDLFTGDISKG 1399
 DB 285 K-----KYNVYSEDLCKDKI-----SLQDFGDIIMGTDM-EG 317
 QY 1400 HGECSKLEQIDSLFKNGDKSPKGTROEWWTEHSEIWEAMCALYIGAKNDPTEN 1459
 DB 318 IGSVLVQNNLRISFGTISAELD--RKKWMDHKADIKAMLSYKE-----KNR 366
 QY 1460 YGVNNVKSSTLTLEFARPOFLRWLTWYDYCYTRQYLYKDVQEKSKNDQKCDT 1519
 DB 367 YSAWCK-EDVQIKVE-----POLYRWIREGRDYMSEFRQRRLKNECE--DKLYST 418
 QY 1520 -----ECKKCKDYKYKMK--KKEWIPQDKYKDEDRKKRFRQIG--VMYDTY 1567
 DB 419 MLICLTPPCNNACSSYEWETGKKQWDVLTSTKSSVKKQKITENIARAYDILKELN 478
 QY 1568 GTNADTLNKRFTASCGRKPGASVVOIRIOLKQAYYADAKCGCTFIENDKTYNI 1627
 DB 479 GFNEVTFENE-----INKRDLYN---YF-----CYCIVQEARKNQOE 513
 QY 1628 SSKDKCKGLVKEANTGAIKQNKGPNNYNNLKELTEDVLPSPRLRLICFH-----AL 1679
 DB 514 N-----YKNVSG--ESKAPSS--NPINEAVKSSGCGKQVEDSAHRSVNEGSKS 561
 QY 1680 DGNITDEVDENLKRRLMVAATEGYNIGQYKKEKKEKITSDAKTSYVPCSA 1739
 DB 562 STNADPGSPGPGPASVDEKAGVPALSAQGGDKPPAPAAATESAHPASAKTPTTA 621
 QY 1740 MKSYFYDLRIILGIDNLEKQTEENLKKIFPKNGTSYVK-----GSDSTTG 1789
 DB 622 -----TEENKQRTQ--VDSVAGDGDKAPPTYSVSPVSGAGDSRST 662
 QY 1790 PGSTARFENNEFCYWNAMICGKRGDRDNGSNSARSDEDLKCGSVPSD--DYPM 1847
 DB 663 PAS-----HLAGENGE--VHNG-----TDTEPKEDGKADPOKNTIEVKGQOTDRSGSL 711
 QY 1848 GKNDEGATQFLWPAWEGMDFCKHKEKELKYGAC-----NDYICGDMD--KRX 1899
 DB 712 GPHYDER-----ASLGE--THMEKDTETGSLTPPEQVNVSAVDGNGVSGGSK 759
 QY 1900 CTDACTQYKPFISEMKPOYERQIKKYGKGNKDIYSEHPVADADAREYLDKQLKICEN 1959
 DB 760 QNEGATL-----SGAESLSSSESVHTINTHTGL-----EN 792

1779 YKGSSTGNGPGSTARKFFMNNKCYWNNAMICGRGRDNGSNGSARSDEDLKCGS 1838
 DB 655 V6-GKDS---GPTSTASHALAGENG---VANG-----TDEPEDEKDKPPORDKVGK 703
 QY 1839 VPSDD--DYPMGKRDGEGTAYOFLRWFPAWGEDFCFKHEKELEKLGAC-----NDYTGC 1891
 DB 704 QPDDSDSGSLGHTHER-----ATLGE---THMEKDTAGSSTLTPRONVSAS 751
 QY 1892 DNEED---KRRKCTACTOYKKFISEMKPOYKQIKRYGENKDKIYSEHPVAKDAEDAREY 1948
 DB 752 DNGNVGSGNKGEGATL-----SGAESLKNGESVHKTIIDMTYHG 792
 QY 1949 LKQALKICEKSGDEYCKMKDVSTORLTDGNSQWMPASLDEPEVEGCKCQVPRG 2008
 DB 793 L-----ENKNGNE---KDPKRDHFNMMNDLQASDHTSD-----827
 QY 2009 PVRRETPSPRVAS---LISATASKKEAKTAPPTKOPKRYENLTTEKRAQTR---RA 2061
 DB 828 -----QTSDDHTSSDQTSDDHTSSDQTSDDQTSDDQTSDDQTSDDQTSDDQTSDD 882
 QY 2062 AQOTR---RSTATTES-----DV-GTWKALIS---NKPSRGG 2096
 DB 883 SEDMSKGFERNNSNELYSHNNLNKLRDQYEHHDYKATREKILIMSEVKKCNRAS 942
 QY 2097 IECNPKTYGQYKMGCIYKSKENE--NGICMPRRKRL-----CINNIQIYLYETE 2147
 DB 943 VKYCN-----TLEDKMLSTCSRRERKNCISDSCMLYFELYEYEF 986
 QY 2148 NKDNDIKKAFIC-----AIEFQFLMKYIEN-PAENELONGTIPDEF 2193
 DB 987 NCKKKEFEDPSYECTFGKSGTGYVATGAGFLIILLFASMAASNDYEEATPDEF 1044

RESULT 11
 MST2.DROHY
 ID MST2.DROHY STANDARD: PRT: 1391 AA.
 AC 008696:
 DT 01-FEB-1995 (rel. 31, Last sequence update)
 DT 01-FEB-1995 (rel. 31, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Axoneme-associated protein mst101(2).
 GN mst101(2).
 OS Drosophila hydei (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7224;
 RN [1]
 SEQUENCE FROM N.A., AND CHARACTERIZATION.
 MEDLINE=95045538; PubMed=7957199;
 Neesen J., Padmanabhan S., Buemann H.;
 "Tandemly arranged repeats of a novel highly charged 16-amino-acid
 motif representing the major component of the sperm-tail-specific
 axoneme-associated protein family Dmst101 form extended
 alpha-helical rods within the extremely elongated spermatozoa of
 Drosophila hydei.";
 RL Eur. J. Biochem. 225:1089-1095(1994).
 CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY
 CC SPERMATIDS.
 CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
 CC -1- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
 CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
 CC REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: X73481; CAA51876.1; -
 DR PIR: S34154; S34154.
 DR Flybase: Fggn020733; Dmst101(2).
 KW Sperm; Repeat; Multigene family; Polymorphism.
 FT DOMAIN 332 1268 59 x 16 AA APPROXIMATE TANDEM REPEATS OF
 FT SEQUENCE 1391 AA; 159000 MW; 1B2A568F30F48878 CRC64;
 Query Match 1.6%; Score 317; DB 1; Length 1391;
 Best local similarity 19.2%; Pred. No. 2.5e-07;
 Matches 334; Conservative 232; Mismatches 660; Indels 510; Gaps 75;
 QY 1507 ECKSDNOLKCD---TECNKK---EDYKMKKKKEMIPDQKYYKDER-----DKRR 1553
 DB 50 ESTDDSNQIRCDADTAADKAKKOKKEEDQAKIREYKRECLKQKHYLAEEIRSGEKDR 109
 QY 1554 FDRQHIGVAVTDYTGNNADYLNKFTASGQDP-----GSASVQNRNOLKQ-QAYVD 1607
 DB 110 ILIQQMKCLTDGMRKACTIARAKLIA--DKELAVQCAALSKDKVALAKKCEERS 166
 QY 1608 ADKRCGCTFEIENDRYTNISKDC-KGLVKEANGAIKMOKGPNYNNLKELEEDYL 1666
 DB 167 KEKECQNSPABD-----KDRTKGKTGKSGSGKRRKSTENRAKKGKLYKNP 218
 QY 1667 FPSRLRLICFHA-----LDGNYTDP-----VKDENGILRRRLM-EVATEG 1706
 DB 219 --TQKLEHCISQWADCECRONTEDERRRLAASYKCMGTIKSICRRVIAEMCEAG 276
 QY 1707 YNIGQYKKEKKEKIKTSDAHKYSYEVPPCSAMKSPFDRIILIGIDNLEDE--KQRT 1764
 DB 277 YKNS---SEPKKKKKKKKDEKKEK-----ELEREILKEQA 309
 QY 1765 EENLKRIFNKNGTSYKGSDDSTGNGPGSTARKFFMNNKCYWNNAMICGRGRDNGSNG 1824
 DB 310 EEEAK-----IRGVKKEVKKKCKEAL--KKKKD---LG 339
 QY 1825 NSARSDKLLKCGSVSDDYPMGKNRDEGTAYOFLRWFPAWGEDFCFKHEKELEKLVGA 1884
 DB 340 RKKEEAERKKCAALAK-----KQKEDEK---A 366
 QY 1885 CNDYTGDNF-DKRRKCTACTOYKKFISEMKPOYKQIKRYGENKDKIYSEHPVAKDAE 1943
 DB 367 CKELAKKKKKEADKKCKEEAANKERK--AAEKKKCAAEKKEAERKKEAARKE 424
 QY 1944 DAR-----EYLDKOLKICEKSGDEYCKMKDVSTORLTDGNSQWMPASLDEPEVEGK 1999
 DB 425 AAERKKCEELAKKIKKAERKKCEAARKEAARKEAARKEAARKEAARKEAARKEAARKEA 478
 QY 2000 CNCQVPRGPRVRRTPSPRVSLISKATAS--KKEAKTAPPTKOPKRYENLTTEKRAQT 2056
 DB 479 CEETAKKGGKGVARRKCEELAKKIKKAERKKCKKLAKKEKEAEKKEKCAAKKKEA 538
 QY 2057 RTSRADQPTKRTSTATTESDVGTVWKAALSKPPSRGIECNPKTYGQYKMGCIY 2116
 DB 539 EKKCKEAAKKEEA-----EKKCKES 562
 QY 2117 KSEKENGICMPRRKRLCINNIQIYLYETENKRDNDKEAFIKCAIETQFLMKYIE 2176
 DB 563 AKRKE-----AAEKKK-----EKAERKEAARKEAARKEAARKEAARKEAARKEA 593
 QY 2177 NPAENELONGTIPDEFKRIYTYDYDKDMFFGTDISNDKTIYTVNSVYTLINENNK 2236
 DB 594 --AAKE-----KEVAERKKCEELAKKI--KK 616
 QY 2237 KQKKKDEELAKTFEMKNNKFTIEMGIYGLTYLTPENR-----EKIDNYQNDMTL 2291
 DB 617 AAERKKCKEAK--KEK-----AAERKCGELAKKIKKAERKKCKKL 658
 QY 2292 TPLSEFVKRQPLRWFTEWAEFCNKRKEOLK---LEAGCKEYECNGSNDKTOCBAC 2348
 DB 659 AKKEKEFAEKK-----CEAARKKRKAARKKCAEAARKEAARKEAARKEAARKEAARKEA 705

QY 2349 ACVTYONFIKKKTEVEROR-----EKKKDKGKKYKDYSPERDIEKATCAHEYLNMK 2403
 DB 706 AA-----KKEEAERKCEELAKKIKKAEEKKKCKLAKKAGKGE-----NK 749
 QY 2404 LKEICGN-----KDCSCMQKPSOLPK-----TTOOSSSANDMPESLDVPEE 2448
 DB 750 LKR--GKKKKKALLKKEKKCKRELAKKAEEKKKCKEAKKEKKAEEKKCKEAKKREK 807
 QY 2449 FNCCEPELSKSGSMITHKTEPKIPMNCVEKAAYLSKSEANNMDITLKEFIPTEST 2508
 DB 808 AEKKCKEETAKK-----KREAAKK--KC-EKAARKREAE-----KKR--CERT 848
 QY 2509 KEKESKNSWTNNPCPKPYADKTYGRNPNCEBNEBNEFKYDKYKSKSYOE-- 2566
 DB 849 AKRKETA--EKKKCE--KAARKKQAAEKKCKEAKKREKKAEEKKCAEAAKKEKELA 904
 QY 2567 KRRVCPVPRREHMLNLEIKIERLKDSNYLLKAVRRTARANGIDILINPSENCAN 2626
 DB 905 EKKKCEBAKK-----EKVAAEKKCKEELAKKIKKAEEKK----- 940
 QY 2627 PICDINKYSPADIGDIYRGDTMLRIGYLPVYEIKLYKFEYLYGKWRNKNKGRNRYNDY 2686
 DB 941 --CKKLAKKREKAGE-----KNKLKK--AGKGRKKCKKLKSKSR 977
 QY 2687 QFRSAMDANKRDIWKAATCAPEDAKLFKRG--RMOGFERITLIDCKGKHD----- 2738
 DB 978 AAEKKCAIAKKEKAAKKAEEKKCEERAKKQCEERAKKLEKAEAOKECEERAK 1037
 QY 2739 --DPPVDYIPQFRWMTSEYCKKALMEELKFKSCDCKTSDRCNDYDENCKQ 2795
 DB 1038 KLEKAAEKKQCEERAKKLEKAEAO--KQCEERAKKLEKAAEKKQCEERAKKLEKAEAO 1095
 QY 2796 CTRCOEYNEFLKMSLFDIOSNKYKELYEOPITYTKISTYDHYONFYOKLTFRSECSV 2855
 DB 1096 CEERAKKLEKAAEKKQCE--EERAKKEKAEKKQCEERAKK-----EKEAAEKKKCAE 1146
 QY 2856 EEFSE-YLHETSKLNYKFNENDGSSNIRTYAFEEETPKSYKAC--SOTLPSKPLDNC 2911
 DB 1147 AAKKEKATEKCAEAAKKEKAAEKKKCAE--ABAARKEKAAOKKCAADIAKKEOE-- 1201
 QY 2912 PPDONKDGCKELQTFPESKNDYNNLNMWNYLVINSDDNKGVLIPRRHLCRPT 2971
 DB 1202 PAEMK--CEE-----AAKKEK-----AAEKQKAKA-- 1227
 QY 2972 AVNYRGKELIKKLLKLSAFSGOLGOKYSEELCEPAKYSYADYSIITKTDMD 3031
 DB 1228 ---AKKEKAAEKKKCAEAAKKEKAAEKKKCAE--AAKKEKAAEKKKCEKAEKRA- 1278
 QY 3032 TSLSEKIKKIEFTSEATENRKTWNENNRQITWAMLCGYKLTATSGVTLDGWOQLPDE 3091
 DB 1279 AALKQCAKLVIRAKKALRRK-----CAIIAKKAKKAAAEKKEC-----E 1318
 QY 3092 ETNOGLRLILIAKQACKEKKHVSLSLTKCPRSNEDNFEASELLRQPCONDYRK 3147
 DB 1319 KLAKKAKAEIEM--KKCAKLAKKKREAEKKKCAKLAKKKEKAAE--KKRCK-DLAK 1370
 RESULT 12
 MYSLDICTD STANDARD; PRT; 2116 AA.
 AC P08799;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-OCT-1989 (Rel. 12, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Myosin II heavy chain, non muscle.
 GN MHA.
 OS Dictyostelium discoideum (Slime mold).
 CC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 CX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87092266; PubMed=3540939;
 RA Martick H.M., de Lozanne A., Levinwand L.A., Spudich J.A.;

RT "Conserved protein domains in a myosin heavy chain gene from
 Dictyostelium discoideum.";
 RT Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
 RN [2]
 RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
 RC STRAIN-AX2;
 RX MEDLINE=90353583; PubMed=2387408;
 RA Lucke-Vielmeier D., Schleicher M., Grabatin B., Wampler J.,
 RA Gerisch G.;
 RT "Replacement of threonine residues by serine and alanine in a
 phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
 RT FEBS Lett. 269:239-243(1990).
 RN [3]
 RP PHOSPHORYLATION SITES.
 RX MEDLINE=88112226; PubMed=2828113;
 RA Wagle G., Noegel A., Schaefer J., Gerisch G.;
 RT "Phosphorylation of threonine residues on cloned fragments of the
 Dictyostelium myosin heavy chain.";
 RT FEBS Lett. 227:71-75(1988).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
 RX MEDLINE=95345066; PubMed=7619795;
 RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M.,
 RA Rayment I.;
 RT "X-ray structures of the myosin motor domain of Dictyostelium
 discoidium complexed with MgADP, BeFx and MgADP.ALFA-.";
 RT Biochemistry 34:8960-8972(1995).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
 RX MEDLINE=95345067; PubMed=7619796;
 RA Smith C.A., Rayment I.;
 RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
 truncated head of Dictyostelium discoidium myosin to 2.7-A
 resolution.";
 RT Biochemistry 34:8973-8981(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
 RX MEDLINE=96206189; PubMed=8611530;
 RA Smith C.A., Rayment I.;
 RT "X-ray structure of the magnesium(II) ADP vanadate complex of the
 Dictyostelium discoidium myosin motor domain to 1.9-A resolution.";
 RT Biochemistry 35:5404-5417(1996).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
 RX MEDLINE=97452580; PubMed=9305951;
 RA Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
 RT "X-ray structures of the MgADP, MgATPgmas, and MgMPPNP complexes
 of the Dictyostelium discoidium myosin motor domain.";
 RT Biochemistry 36:11619-11628(1997).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
 RX MEDLINE=98070605; PubMed=9405148;
 RA Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
 RT "X-ray crystal structure and solution fluorescence characterization
 of Mg₂(3')-O-(N-methylanthraniloyl) nucleotides bound to the
 Dictyostelium discoidium myosin motor domain.";
 RT J. Mol. Biol. 274:394-407(1997).
 RN [9]
 RP FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
 ACTIVITY THAT IS ACTIVATED BY ACTIN.
 CC SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
 INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
 LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
 (MLC-2).
 CC SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
 CORTEX.
 CC DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN BE FURTHER
 SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 CC DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES

CC THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE
CC ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
CC POSITION (688).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IO DOMAIN.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL; M14628; AAA33227.1; -
DR PIR; A26655; A26655.
DR PIR; S00250; S00250.
DR PDB; 1MMA; 03-DEC-97.
DR PDB; 1MMD; 17-AUG-96.
DR PDB; 1MMG; 03-DEC-97.
DR PDB; 1MMN; 03-DEC-97.
DR PDB; 1MND; 17-AUG-96.
DR PDB; 1MNE; 17-AUG-96.
DR PDB; 1YOM; 23-DEC-96.
DR PDB; 1LYK; 28-JAN-98.
DR Dictydb; DD01008; mhca.
DR InterPro; IPR000048; IO_region.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
KW Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
FT DOMAIN 1 761
FT DOMAIN 762 791
FT DOMAIN 817 2116
FT NP_BIND 179 186
FT DOMAIN 638 660
FT DOMAIN 738 752
FT MOD_RES 130 130
FT MOD_RES 678 678
FT MOD_RES 1823 1823
FT MOD_RES 1833 1833
FT MOD_RES 2029 2029
SQ SOURCE 2116 AA; 243871 MW; 2PC3770BB1E556A1 CRC64;

Query Match 1.6%; Score 313.5; DB 1; Length 2116;
Best Local Similarity 18.0%; Pred. No. 6.3e-07;
Matches 377; Conservative 336; Mismatches 827; Indels 551; Gaps 81;

QY 420 NINSEYKQEY---KLEFQY---ATNDTFLNLNEGKCYCKGGLGENDIYFTNSADD 472
Db NINSEYKQEYFNHMFLEQEEYLEKINWTFIDE-----GLDSQATIDIDRQRP 522
QY 473 KGIFYREYQVCPDCCVKDGIKTYTHKSDNDRVNNEDYKPPWGVKFNITVLYSGNE 532
Db KGIFYREYQVCPDCCVKDGIKTYTHKSDNDRVNNEDYKPPWGVKFNITVLYSGNE 532
QY 523 PGIALLDEQSVFPA---TDNLTLLKLSHFSSK--NAYEERFSKTEGVTYHAGOV 577
Db PGIALLDEQSVFPA---TDNLTLLKLSHFSSK--NAYEERFSKTEGVTYHAGOV 577
QY 533 QGDITQLENCNSSTNYKDKNNQKWCYKDEINRCKLEQNTFENNDPKIIS-----F 588
Db QGDITQLENCNSSTNYKDKNNQKWCYKDEINRCKLEQNTFENNDPKIIS-----F 588
QY 578 MYEQDPLE-----KNRPLQDDLELCRQSSDNVY-----TKLND-PNLSRAKKG 624
Db MYEQDPLE-----KNRPLQDDLELCRQSSDNVY-----TKLND-PNLSRAKKG 624
QY 589 HNFELVWYLLRDITKIMNDLKTKCINNTTTH---CIDEENRNCLEPDRWVKQKEEWN 645
Db HNFELVWYLLRDITKIMNDLKTKCINNTTTH---CIDEENRNCLEPDRWVKQKEEWN 645
QY 625 ANFTVAAY--KQOLA---SIMATLETJNPHEVRCIIPNNKQJPA----- 665
Db ANFTVAAY--KQOLA---SIMATLETJNPHEVRCIIPNNKQJPA----- 665

QY 646 IKKLFTRKNNIQOSYSYNNNNLEBGFYFVKMDKLDDEAKWELMNRKKNEFSLN 705
Db IKKLFTRKNNIQOSYSYNNNNLEBGFYFVKMDKLDDEAKWELMNRKKNEFSLN 705
QY 666 -----KLEDYVLDQLCNGLEBGRITRRGFPNRI 697
Db -----KLEDYVLDQLCNGLEBGRITRRGFPNRI 697
QY 706 NRDYLENAIELLDHLKETATICKDNNTNEACETSHATNPCKPGRGQPRKNKEIA 765
Db NRDYLENAIELLDHLKETATICKDNNTNEACETSHATNPCKPGRGQPRKNKEIA 765
QY 766 QYERKSA---YEARNRGLHKLKGAHEGIRKGRKDKFKN-----LCRIMIKHSR 816
Db QYERKSA---YEARNRGLHKLKGAHEGIRKGRKDKFKN-----LCRIMIKHSR 816
QY 743 KIFPRAQOLARIBEARQRISEII-KAIOATGWIMARKYKQAREHTVAARIIOQLRA 801
Db KIFPRAQOLARIBEARQRISEII-KAIOATGWIMARKYKQAREHTVAARIIOQLRA 801
QY 817 NLGFSNCPGCKGTGDSIQTRFYVGTW-----EVDPEHKKKHEDYIMPPRRRH-CTS 870
Db NLGFSNCPGCKGTGDSIQTRFYVGTW-----EVDPEHKKKHEDYIMPPRRRH-CTS 870
QY 802 YIDKSWP-----WMLFSAFRLKRRNEKEIKEREIELELS 842
Db YIDKSWP-----WMLFSAFRLKRRNEKEIKEREIELELS 842
QY 871 NLEHLQDDBPLNGNIYVDLVNNSFLGDVLLSAKYENKTIIRYKKNNKKGK----- 924
Db NLEHLQDDBPLNGNIYVDLVNNSFLGDVLLSAKYENKTIIRYKKNNKKGK----- 924
QY 843 NLDSITQKRLKSLND---TESNVLDLQRLKAEETLKAMTDSKDLAEQKRELEIR 899
Db NLDSITQKRLKSLND---TESNVLDLQRLKAEETLKAMTDSKDLAEQKRELEIR 899
QY 925 -----EVDPKHQTTICRAIRYSFADIGIIRGRDLERNRDMVKLQGHLETVFGNIHK 978
Db -----EVDPKHQTTICRAIRYSFADIGIIRGRDLERNRDMVKLQGHLETVFGNIHK 978
QY 900 VEDMESELDKRLALENIQOKRSVEE-----KYRDLSEELQERQKRLNLEKTKKYE 954
Db VEDMESELDKRLALENIQOKRSVEE-----KYRDLSEELQERQKRLNLEKTKKYE 954
QY 979 SLK--GKNDKYNDDAKYTKLKRENNWEANRAKYWEAMKCDIYTKKSGHSTQSYCG 1036
Db SLK--GKNDKYNDDAKYTKLKRENNWEANRAKYWEAMKCDIYTKKSGHSTQSYCG 1036
QY 955 ELEEMKRVNDGOSDTISRLERKIDEL---QKEVEELTESFSESKDKGVLEKTRVAL- 1008
Db ELEEMKRVNDGOSDTISRLERKIDEL---QKEVEELTESFSESKDKGVLEKTRVAL- 1008
QY 1037 YSDHTPLDVIYIPQRLRMTMAEWYCKVOKKEVDKLEKKECKDKONGGCGRRESGTC 1096
Db YSDHTPLDVIYIPQRLRMTMAEWYCKVOKKEVDKLEKKECKDKONGGCGRRESGTC 1096
QY 1009 ---QSELDL---TVRLDSETKDSKSELLROK--KLEELKQVE-----ALAAATPA 1053
Db ---QSELDL---TVRLDSETKDSKSELLROK--KLEELKQVE-----ALAAATPA 1053
QY 1097 TKCTEACNEYNDIIGLKEQWNIISDKYKELHQAQNSV--SNSGIEASSTAKNHDNRVI 1155
Db TKCTEACNEYNDIIGLKEQWNIISDKYKELHQAQNSV--SNSGIEASSTAKNHDNRVI 1155
QY 1054 KLAQEAANKR-----LOGEYTELENERFSEVARSVRS-----KTLLESQVL 1097
Db KLAQEAANKR-----LOGEYTELENERFSEVARSVRS-----KTLLESQVL 1097
QY 1156 EPLSELYQONGGSKNSGTSDESAVITNTTYEVNAGVLYHDTGNFDDCOSQNEPCDEKSD 1215
Db EPLSELYQONGGSKNSGTSDESAVITNTTYEVNAGVLYHDTGNFDDCOSQNEPCDEKSD 1215
QY 1098 AVNNLEDEE---KNRDLERKKRAL--DAMLEEMKQDLSTG---BKSLYLDKVK 1147
Db AVNNLEDEE---KNRDLERKKRAL--DAMLEEMKQDLSTG---BKSLYLDKVK 1147
QY 1216 GKDNKAYFRDKPDHOGACGCKSGSKPT---RVQ-----IKTKKAEK 1257
Db GKDNKAYFRDKPDHOGACGCKSGSKPT---RVQ-----IKTKKAEK 1257
QY 1148 -QESDMALNRQISELOSTAKLEKIKSTLEGEYARLQGLELEAQLAKSVEKQKKVEL 1206
Db -QESDMALNRQISELOSTAKLEKIKSTLEGEYARLQGLELEAQLAKSVEKQKKVEL 1206
QY 1258 DTECKTVNDILKENDGKQYEDCHPKKNSNGYPMDCGNINLYEDPVCAPPRQKLCVH 1317
Db DTECKTVNDILKENDGKQYEDCHPKKNSNGYPMDCGNINLYEDPVCAPPRQKLCVH 1317
QY 1207 DLEDKSAQ--LAETLAKQALDKLKK----- 1231
Db DLEDKSAQ--LAETLAKQALDKLKK----- 1231
QY 1318 FLANDNEIKKQSVNLKEAFISAALETFSWYYSKSGEGNELDKL-BGKIPPAF 1376
Db FLANDNEIKKQSVNLKEAFISAALETFSWYYSKSGEGNELDKL-BGKIPPAF 1376
QY 1232 ---LEQELSEVQTO--LSEANNKKNVNSDS---TNRHLETSFNMLKLELEAQKAKQAL 1281
Db ---LEQELSEVQTO--LSEANNKKNVNSDS---TNRHLETSFNMLKLELEAQKAKQAL 1281
QY 1377 LRSWFTYFGDYRDLFPTDISKGHSGSKLKEQIDSLFKKNGDKSPKGRQEWMTESH 1436
Db LRSWFTYFGDYRDLFPTDISKGHSGSKLKEQIDSLFKKNGDKSPKGRQEWMTESH 1436
QY 1282 EK-----KRLGLESELKHVNEQLPEEKQKQSEKRRVDEKE--- 1319
Db EK-----KRLGLESELKHVNEQLPEEKQKQSEKRRVDEKE--- 1319
QY 1437 EIWEMALCALVKGAKKDEFTENYNNVRFSDKSTLLLEFAKRPQRLMTEVYDYCY 1496
Db EIWEMALCALVKGAKKDEFTENYNNVRFSDKSTLLLEFAKRPQRLMTEVYDYCY 1496
QY 1320 -----VSELKQLEE-----EVAASKAVTEAKNNKSESELDKIQYADVVS 1360
Db -----VSELKQLEE-----EVAASKAVTEAKNNKSESELDKIQYADVVS 1360
QY 1497 TRQKYLNDVOEKSSNQQLCDTECNKCKEDYKYMKKKEMIPQD--KYKDERDKRRF 1554
Db TRQKYLNDVOEKSSNQQLCDTECNKCKEDYKYMKKKEMIPQD--KYKDERDKRRF 1554
QY 1361 SRSKSVQKLTQAKNLELRLTALEAGOLDPRRSRKKKEFDELEAVKMLEETAKVK 1420
Db SRSKSVQKLTQAKNLELRLTALEAGOLDPRRSRKKKEFDELEAVKMLEETAKVK 1420
QY 1555 DROHIGWATVDYGT---NATDYLRKFTASGDRPGSASVQVRNIOLE--KQAVYDA 1608
Db DROHIGWATVDYGT---NATDYLRKFTASGDRPGSASVQVRNIOLE--KQAVYDA 1608
QY 1421 AERKAKAELEIDYSTKSELDDAKVSSQY-----VQIKRLEELSELRSVLEEA 1470
Db AERKAKAELEIDYSTKSELDDAKVSSQY-----VQIKRLEELSELRSVLEEA 1470
QY 1609 DKHCGKTRFIENDKYNNISKDKCKGLVKEANTGAIKMNKGNPNYNNLEKTEDEVLP 1668
Db DKHCGKTRFIENDKYNNISKDKCKGLVKEANTGAIKMNKGNPNYNNLEKTEDEVLP 1668
QY 1471 DERCNESA-----IKAKTAESALELKEIDDAANNAKAKARRSKSELEIVRAVELEESLEDK 1526
Db DERCNESA-----IKAKTAESALELKEIDDAANNAKAKARRSKSELEIVRAVELEESLEDK 1526
QY 1669 SRRLRICE-----HALDGNITDPEVK---DENGRLKRLMEVAALBGRNLGQYK 1714
Db SRRLRICE-----HALDGNITDPEVK---DENGRLKRLMEVAALBGRNLGQYK 1714

```

Db 1527 SGTVAVEIRKKAIDDLRLADRETESRIKSDKKNRRKQADLEA-----K 1576
QY 1715 EKKKEKIKTSDAKAYSVEPPCSAMKSYDLDLIDLDNLEDEQKTEENKIKTFNK 1774
Db 1577 VEEQREVITIDIRKAKLE-----SDIILSTQDLETTRSHIKTEKSKK--- 1622
QY 1775 NGTSVSGSDSTGPNPGSTARKPFWMENKRCVNMAMICGYKRGHDNSGNSARSDEDLK 1834
Db 1623 -----EQTILAE-----RRAAEBSKRAADEIR 1645
QY 1835 KCGSVPSDDVPMCKNDEGTAYOFLWFAEMGEDPCKHKEKEKELVGCANDYTCGDNE 1894
Db 1646 K-----QVMOEYDE-----LEAQLDSERAALNASEKIKSLV-----AEVD 1681
QY 1895 DKRRKCTACTOYKFKFISEKMPYOKOIKRYGENKDKIYSEHVADAEDAREYLDKOLK 1954
Db 1682 EYKQLEDEITAKDLKVA-KRALVELE---EYRQOLEEEDSRSLIEDSKRLITTEVE 1737
QY 1955 KIGENSGDEYKCMQDVSTQRLTDGNSQMPASLDDEPEVEGKCNQVPRPRVRE 2014
Db 1738 DKKKDAEVEQNTKLDKAKKRLTD-DVDTLKRLQLEDEKKK----- 1777
QY 2015 TSPRVSLSKATASKKKEATAPPTQPKKVENLTTEMRAGOTRRRAAOOTRRK----- 2068
Db 1778 -----INESERAKKRLS-----ENEDFLAKLIDEVANRSRAEDRKRYEKDLK 1821
QY 2069 -----TSTATTESDVTGTVKAILSNKPNDSRGIEGCPKTYGQYPRKWCIVGSKEN 2121
Db 1822 DPKYKLNDEAAATQTEIGA---AKLEQIDE-----LRSLQLEB 1858
QY 2122 ENGICAPPRKRLCINNIOYANETENKRDNDIKFAFK-CAAIETQFLMKYLIENPAA 2180
Db 1859 QAKATPADSKTLBEELNLRQIED--EGKIKMLEKELKALBELLELREYVE--A 1914
QY 2181 ENELQNGTIDPEKFRIMTYTGDKMPFGTIDISNDKIIITVNSVTTLINNNKKKODK 2240
Db 1915 EDKSEA---EQSKRLVELELDAR-----RMLQKEIDAKLEAEDAKSNLQEIYEAKEK 1966
QY 2241 KQDEELRKIFWEKKNKFTWEGMITYGILTHLDNEKEKIRDNQYNDMTKLTPLSEEPK 2300
Db 1967 LEEBESIAKTNDSRSKRL-EAELIDALTAQV-DAEQKAK---NQJIKENKIKTELKEK 2021
QY 2301 RPOFLWETEMAEFCNKRKEQLL--KLEAGKEYEONGSNGKQOECACVYONFK 2358
Db 2022 K-----FGSEKTKTKELVYEKLETDYKRAKKEADQOOR-----LTVENDLR 2066
QY 2359 KKKTEYEROREKFKDKDQKTKYKDYSPSTERDIEKATCAHEYLNMKIKELCG 2409
Db 2067 KHLSEI-----SLKDAIDKLDORDHDKTRKLELETWASKIEMQRKADPFG 2112

```

```

RL Mol. Gen. Genet. 230:277-287(1991).
CC -1- FUNCTION: CONTROLS NUCLEAR MIGRATION. NDM1 SPECIFICALLY CONTROLS
CC THE INTERACTION OF THE BUD NECK CYTOSKELETON WITH THE PRE-
CC DIVISIONAL G2 NUCLEUS PERHAPS BY RECOGNIZING G2-SPECIFIC
CC CYTOPLASMIC MICROTUBULI OR OTHER COMPONENTS OF THE NUCLEAR
CC ENVELOPE.
CC -1- MISCELLANEOUS: ADDITIONAL REGIONS OF LOWER HOMOLOGY TO THE REPEAT
CC CONSENSUS (ALWAYS STARTING WITH PROLINE) ARE FOUND IN BOTH
CC FLANKING DOMAINS OF THE TANDEM REPEATS.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: X61236; CA43554.1; -.
DR PIR: S19052; S19052.
DR SGD: S0002557; NDM1.
DR InterPro: IPR001849; PH.
DR Pfam: PF00169; PH; 1.
DR SMART: SM00233; PH; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR KW
FT REPEAT.
FT DOMAIN
FT 593 1384
FT REPEAT 593 656
FT REPEAT 657 727
FT REPEAT 728 798
FT REPEAT 799 862
FT REPEAT 863 926
FT REPEAT 927 990
FT REPEAT 991 1054
FT REPEAT 1055 1118
FT REPEAT 1119 1182
FT REPEAT 1183 1246
FT REPEAT 1247 1310
FT REPEAT 1311 1374
FT REPEAT 1375 1384
FT DOMAIN 2573 2683
SQ SEQUENCE 2748 AA; 313202 MW; B2FBD67C9F6211AE CRC64;

Query Match 1.68; Score 309.5; DB 1; Length 2748;
Best Local Similarity 17.98; Pred. No. 1.3e-06;
Matches 600; Conservative 467; Mismatches 1122; Indels 1161; Gaps 164;

QY 95 HTNLRVDVNLRHPCRGREONRDE--DEES-----ECGKIRNYKRRND 137
Db 3 HNN-RHKKNDDKSSAGQYANSIDNSLSQSVSTNGVTRMANLKADCGSGDGSDTKRF 61
QY 138 AIACAPRRRHMCNKLLEALNDITQNIHDLGNVLTARYEGESIVNHPHKGTSIDACT 197
Db 62 SISSILSKR-----ETKDVLPREFAGSSSH-----NGVLTANSKDMFTLESENLLVEER 112
QY 198 ALARSFADIDIVAGIDIMFKPNVHDVE--TGLREYFKIHDQMEDEYKNDYDPDSGAY 255
Db 113 KLGSSNNAKNEQIKSLKQIKESISDKIETLITNOKSKSMELDSTKO----- 158
QY 256 YKLEAMNWNVRNKKVWEALTCDASYSKGYFMQSESNTPLESNPKCGHKQKQPTNLDYVP 315
Db 159 -----LN-----MDL-----ESKL-TNLSMPC 174
QY 316 OYLRFWDE-----WGEFPCRRKRIKIKKYKDCGRNDEKRYCSHNG-----HDC 359
Db 175 KQLEKLEKTKESKWNDE--KESLKLK-----TDLELITLTKKGMENDLSSOKLHYDK 225
QY 360 TTYINRKGILHLDNK-----CTDCSTCKYFVWVIGNOEAFFKQKREYKQSYLS 412
Db 226 EISELERITLDNNENDRLISVSDLTSEINSIQ---SKRTERIKIQKO-----LD 273

```

RT highly repetitive 313 kDa NDM1 protein.";

DR EMBL: AL096710: CAB63786.1: -
 DR EMBL: AL096710: CAB63787.1: -
 DR EMBL: L11690: AAA52288.1: -
 DR EMBL: M63618: AAA53606.1: -
 DR EMBL: U04850: AAA57184.1: -
 DR EMBL: X58677: CAA1528.1: -
 DR EMBL: M2942: AAA5358.1: -
 DR Genew: HGNC:1090: BPAG1.
 DR MIM: 113810: -
 DR InterPro: IPR001101: Plectin_repeat.
 DR InterPro: IPR001452: SH3.
 DR InterPro: IPR002017: Spectrin.
 DR Pfam: PF00018: SH3; 1.
 DR Pfam: PF00435: spectrin; 3.
 DR Pfam: PF00681: plectrin; 6.
 DR SMART: SM00250: PLEC; 11.
 DR SMART: SM00150: SPEC; 3.
 DR PROSITE: PS50002: SH3; FALSE_NEG.
 DR Antigen: Coiled coil; Repeat; SH3 domain; Structural protein;
 DR Cytoskeleton; Cell adhesion; Alternative splicing.
 DR DOMAIN 1 1199
 FT DOMAIN 1200 ?
 FT REPEAT 264 2704
 FT REPEAT 366 449
 FT REPEAT 474 567
 FT DOMAIN 581 633
 FT REPEAT 975 1071
 FT DOMAIN 1179 1380
 FT REPEAT 2165 2222
 FT REPEAT 2223 2260
 FT REPEAT 2261 2296
 FT REPEAT 2297 2334
 FT REPEAT 2385 2421
 FT REPEAT 2492 2531
 FT REPEAT 2532 2569
 FT REPEAT 2570 2607
 FT REPEAT 2608 2645
 FT REPEAT 2646 2683
 FT VARSPLIC 1 55
 FT VARSPLIC 363 367
 FT VARSPLIC 457 468
 FT VARSPLIC 688 692
 FT VARSPLIC 867 884
 FT VARSPLIC 1147 1161
 FT VARSPLIC 2427 2440
 FT VARSPLIC 2441 2704
 FT CONFLICT 1998 1998
 FT SEQUENCE 2704 AA: 313084 MW: A7219E687A634A77 CRC64;
 Query Match 1.6%; Score 309; DB 1: Length 2704;
 Best Local Similarity 17.9%; Pred. No. 1.4e-06;
 Matches 559; Conservative 427; Mismatches 1073; Indels 928; Gaps 140;
 DB 814 SRRNLGFSNG-----PCDGGKGTGDIQTRFVVGTEWEVDPEHMRKHEDVIMPPR 863
 DB 207 SWSGSLSSGMSRLTPSVTPAYTPGPSPGLVNFSSG---YEPNSL---QTLMQI 258
 DB 864 RRICTSNLEHLOTDPHPLNGNIVDVLVNSFGDVLISAKYIIRKTKKNNLKCP 923
 DB 259 RKPLKLSL-----LDONLTEREINNKFFVODLL----- 286
 DB 924 KEVTDPHQTTICRAIRYSFADIGDIIRGRDLERNQDGVKLGHELVFVGNLHK----- 978
 DB 287 -----KNVDEMOYOLDRTKM-----GSDLPVSFVSHLEN-HKNVHRALEEF 325

QY 979 --SLKRGNDKYNDAPKYLKIRE--NWEANRAKYWEAMKCDIKYLDKSGHQSOTSSY 1034
 DB 326 ESLLKRAKISEIQMTAPLKVTAEKJLHRLSEYAKLLFSL-----LKNTRNQEHLD- 378
 QY 1035 CGYSDHTPLDDYIPQ--KIRMTENAEMWCYQYKREYDKLAKCKECKDKNGQCCYTE 1091
 DB 379 -----TLHNFVSRAATNELIMNEKEE-----EVAAYWSEHNTIARAKDYHAELME 426
 QY 1092 SGYGTCTKTEACENYNDIIGLMEQWNTISDKYKEJHEQAOQSVSNGIEASTAANHD 1151
 DB 427 -----LDOKEENIKS--VOELAEOLL-----ENHPA 451
 QY 1152 RNVEEF-----LSELYOQNGSKNSGTSDESNAVIGTWT 1185
 DB 452 RLTIKVLKLSVWLVVEYAVRAAQOTQMSWILQCCQVEQH-----IKENT 496
 QY 1186 TY-----ENVGAYLHDGNFDCCOSQNEFCDEKS-----DGKDNEKAFKDPQDHIG 1233
 DB 497 AYEPFPNDAKEARDYLK--NLKDAIQRYKSCDRSSIHKLIEDLVQESMEKEELLQYKS 553
 QY 1234 AGCKKSGSKPTROYIKTKKKAEEKDKTECKTVNDI-LKENDGRKQVE-----DCHPKKN 1285
 DB 554 TIANLMGKAKTITQLK-----PRNSDCPLKTSIPKAIQDYQIETTYIKDECVLANN 607
 QY 1286 SNGYPRDMQ-----CGNINLYEDPRVCM---PPRQKLCVHPLAN--DNEIKKL-----QSQ 1331
 DB 608 SH-RAKMKVYISPGNEAMV--PVCFTVPPRKNEMV--DIARIDQYQVNLTLWHESH 661
 QY 1332 VNKEAFKISAAAEFTFSPWYYSKDEGNELDK-----ELKECKIPAPFLRSH 1380
 DB 662 INNK-----SVYSMWYLI-----NEIDRIRASVASYSMILIKTMLGHEHQVYL 705
 QY 1381 FYTFGSDYRDL-----FGTDISKHGESEKLEQDLSFKKNDQKSPNKTKQEW 1431
 DB 706 SNIQSRPEFLSDQSQVFGSDITQLEKEVAVVCQYQOELLKSAEREOESVYNYLY 765
 QY 1432 TEHSH-----EIVEMALCALVKGAKGKDFTEYGVYNNVFSQKSTLEEFARPOFLW 1486
 DB 766 SEVRNRIKLRENEDRLIRIIRPLERDLHEV----- 799
 QY 1487 LIEWYDYCYTRQKYLKDVQEKCSNDQKCDTECNKCCEDY-----VYMKKKE 1537
 DB 800 -----FRTEQEKLEKLELRK--DDELGIT--NKCEFPQAASSVPTLSELN 847
 QY 1538 WIPQDYYVQDERKKRPDRDHISGVMTDYGTVNADYLNKKFLPASGDKGSGSVQGRN 1597
 DB 848 VLDQNNQVYSMSSTIIDKYVAISI--YKGYLISFLRK-----TYNLVKNT 895
 QY 1598 QLEKQAYYDADHCGCTKEIENDDYTNISKDKGGLVEKANTGAIKWQNGPNVYN 1657
 DB 896 QAQEAALVLYKLETC-----EEAVIADKNINENL-----ST 928
 QY 1658 LKELTEDVLPFSRLKICFALDGNNTDPEVKQENGLRKLMEVAALEGYNLCQYKREK 1717
 DB 929 LKQWRESEV--DEKQOVFHAL-----EDELQAKAISDEMEFTYERDLDF--DWHKEKA 978
 QY 1718 EKEKITSQAH-KYSYEVPPCSAMKSYFDLRDIIIGIDNLEDEKQTEENLKKINKNG 1776
 DB 979 DQVVERKQNVHVIDNRRLDLEGISLKYRT--YHPLDMIOQVETQTKT----- 1030
 QY 1777 TSVGKSDSTTGNPGSTARKFEWNEKCEYWNAMIGYKRGGRD-----GNSGNSARS 1831
 DB 1031 -----QENQPEKSKTLATQL--NQQKMLVSEIEM--KQSKMDEQKVAQYSAIVKDY 1079
 QY 1832 DLKCC--GSVPDDDDTPMKANDDEGTAYOFLWPFMEWGEDFCFKHKELEKLVGACNDY 1888
 DB 1080 ELQTTMYRAVYDSQKSPVRRRQSSADLIIQEFM----- 1115
 QY 1889 TCGDNDKRRKCT--DACQYKFTISEMKPYQEKQIKKYGKNDKIYSEHPY--AKDAD 1944
 DB 1116 -----DLRTYALVTLMTQYIKFAGD-----SLKRLKEEVYIINSEHPVNIKELEI 1162

QY 1945 AREYLDKQKICKENKSGDEYKCMKD-VSTQRLTQGNSONMPASIDDEPEVEGKCMQ 2003
 Db 1163 KR-----CKETS---EHGAYSDLRQKATYLENKLTGKISELRMY-----AE 1204
 QY 2004 VPRGPRVREBETSPRVSLSKATASKKREKATAPPTKOPKVENLITE-MRACOTRR--R 2060
 Db 1205 LKKOKSVEBELPRVR-----EAAENELR-----KQGNVADISQKTRASEKQYR 1252
 QY 2061 AAOOTRRTATTTESDVTMYKALISNKPDSRGIGECNPTTYGOYPMGCVKSKKE 2120
 Db 1253 RELETTYREKREARELE---RVROLTEAEAKRAVEE-NLNFRTQLEENFTRTLE 1308
 QY 2121 NENGICMPRRKKICINIOY-----LNYENKRDNDIKAFIKCAIEQIFMLKATIE 2176
 Db 1309 DH-----LKRRLSNDLEQKANKMEELRRKRDE-BELLKLIKOMEDLAFQOYAVE 1361
 QY 2177 NPAENELONGTIPDEFKRLMYTYGDKYKDFGTDISNDKILTYT-----NSVTTLN 2231
 Db 1362 KOLKEKOKTELEARRKITEIYQ-----TCRENALPVCPTTQATSCAAVYGLQ 1409
 QY 2232 ENKKKKOD-----KKDEELKITEKPKKITEGMIYGLTYHLTDENEKER 2278
 Db 1410 EHKQAAEELKQOYDELTAANKRAEODMRELYELN-----ALQLEKTSSEKAR 1459
 QY 2279 IRDNYQNDMTKLTPSLEEFVKRPPFLRWTEMAEEFCNKRREOLK-----LEAGKE 2332
 Db 1460 L-----LKKDLDETNNLTBLCKLELE-----KRDQAEKGSQOLRELGROL 1500
 QY 2333 YECNSNDKQTCACACTYQNFIKKMTETEYEROREKFKKDKGKRYKDYSTERDIEK 2392
 Db 1501 NQTTGAEBAMQASD-----LKKIRNYQLELESINHEK--CK-----LQREVD 1544
 QY 2393 ANCAH-----EYLNMKLELCKGNKDCS-----CMQKP-----SSQLPKTQO-----SOS 2433
 Db 1545 ITRAHVAKNQIHLNSQHSFRDEKELRLDIOCRKSDHLEQEKSEHOLLONIKAEK 1604
 QY 2434 DANDPESIDYVPEEFNKC-ECPELSKSGSMITHKITEPKIPNMCVEKAAVYLSKEAN 2492
 Db 1605 ENNDKIORLN-----EELKEKNECAEMLKQVEBELTRONNETKLMQRIQ-----AEEN 1654
 QY 2493 NMDITLKEKFIPIESTKESKSNMTNNPCPKPIYADKTYIGRRNCEENRENFYD 2552
 Db 1655 ---LYEKO-----TIOQCEALKTQAD-----GPKDOLRSTNEH----- 1686
 QY 2553 YEMKCYKSKFYOEKRYVCPRRHEMCLRN-----LDEIKTERLKDSYILKMYRR 2604
 Db 1687 ---LHKQTKTEQDF-----QRIKICLEEDLAKSOLVSEFK--QKCDQOITITQNTKK 1734
 QY 2605 TARNGIDILIKNFSENGCANPICDITMKYSPADIGD-IVAGTDMLRIGGYLIPVEIKLY 2663
 Db 1735 EVRNINAEI--NASKKEKRRGEOKYOLOAOYQOELNRLKRYODELHLKTEIEQMTIRKM 1792
 QY 2664 KVEFIYIGKWRKNKNGRK-----YNYQUTRSAMWANDKNDIKAMTCAKPE 2711
 Db 1793 VLFQESKGFQSAPEFKKMEKESVITENDISGIRDFVSLQDEN-----SRQOE 1846
 QY 2712 DAKL-----FRGRDGEFRITLIDKCGSHKDDPPVD-----DYIPORFRW 2752
 Db 1847 NAKCETNIKELEROLQYREMOQOGMEANHYQKQCKLEDELIAKREVENIKQKMDQ 1906
 QY 2753 MTEMSEYCKALMELEKFKKSCCHCKTSDBCKANDYENCKEOCKTRCOEY----- 2803
 Db 1907 QIKHEHOLVLLQCEIQK--KSTARDQ-----TFKPDF-----EMVVKECQSHGELSHTNG 1956
 QY 2804 ---KNFVLMKSLFDIOSNKKYELQPIYKISYDHYQVQFKITFKSECVES 2857
 Db 1957 HLHPTPSPLLRWYQ---EPQPLEKMKQHRVQYQI--KEVO--FQPPGAPLEKESKQOC 2009
 QY 2858 FSEYLIHETSKCLANKFENDSS-----NIRYVAFEE--PKSYKE--ACSCYI----- 2902
 Db 2010 YSEYFQSTSEIQTITPETPITRLSEIERIDQALNNSPAPRYQMACEMELVYKILP 2069
 QY 2903 ---ISKNLIDNCPIDQKDGCKE---LQTFPFCKNDYDNNIDN 2940

Db 2070 LEIAKNKOYDMHFEVTLTKQENKPPSA-EEMALECGRASGGLKAGDPLKGLDEPTFON 2128
 QY 2941 KNAVYLVNSDDN-----KGYLIPR-----RRHL 2965
 Db 2129 FGDBHACSVRDEDFKFOGLRHTVATROLVKALDMRTIEQLRLGLKVEEYQTLNKL 2188
 QY 2966 CRRPITAYNKRKGD-----LLKKLITSAFSGQ-----LLGQKYSSE 3006
 Db 2189 TQATSIAGILESTIEKESFASAEIRITIDKVALALAEQAANGFIIDPISGOTYSVED 2248
 QY 3007 EL-----CEEMAK-----YSYADSDIIGCTMDMTSLSEKIKKIFESNKEAT 3050
 Db 2249 AVLKGVDPDEFIRILLEAKAVGYSY-----SKSLSYF-----QAME 2287
 QY 3051 NKTWMENNRQIWMALCGIKIATSKYTLDEGWCQLEPKDETYQFLWLIEMAKQCKE 3110
 Db 2288 NML---DROGKILTA--QIASGVIDPVRGIRVPEIYALQGL-----LNNAILQ 2335
 QY 3111 KHVSDSLKTCPCRSNEDNFA--SELLRQPGCONDRIKYSINILIKNTMELNIRK 3167
 Db 2336 FLHESSMTVFP--NPNNKALYSELIR--CFEDVESOCFLPFGERNISLNVK-- 2389
 QY 3168 QIKQSSGNIDNKP--SEENVOSYIKSDSQALELNDINEIYGTKNNEENFEVLEKTL 3226
 Db 2390 --KTHRISVDTKGSSELYVE-----AFQRLNLEKSIYELSGOQYQWKEAM--- 2435
 QY 3227 YPGLYFEVDETHKNHVLG-----NIKEE-EQTVPRKALFFTPHYVDSFYQAPLESTH 3278
 Db 2436 ---FESYHSSHMLTDTKGLHFNINEAIEGFTIDKALY-----KRYOGLITLT 2483
 QY 3279 RVA-----QYDPKNDILKSSIVYVSLG--LIALHFMKKFKSSVDLIRIINIPQGEY 3331
 Db 2484 ELADSLSLRLVPKKD-LHSPAGYWLTAASGERISVLAASRNLDVRLTALRCLEAOSTG 2542
 QY 3332 GMPLEKSNRRIYRSGPYGKTYIMEGDTSGDEKYMMDLSSDITSESEYEELDIN 3391
 Db 2543 GI-----IDPLGKTKRYVAELRHGLVDEGF-----AOQLROCELYIT 2580
 QY 3392 DIYVGSPPKYKTLIEV 3408
 Db 2581 GIGHPIITNKMMSVEAV 2597

RESULT 15
 ATRX_HUMAN STANDARD: PRT: 2492 AA.
 AC P46100; P51068; Q15886; Q9NNTS3; Q9H0Z1;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE transcriptional regulator ATRX (X-linked helicase II) (X-linked nuclear protein) (XNP) (Znf-HX).
 GN ATRX OR RAD54L OR XH2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANT S-1860, AND
 RP VARIANTS ATR-X.
 RX MEDLINE-97123494; PubMed-8968741;
 RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
 RT Gibbons R.J.;
 RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome.";
 RL Hum. Mol. Genet. 5:1899-1907(1996).
 RP [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
 RX MEDLINE-97386582; PubMed-9244431;
 RA Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,
 RA Colliaux L., Schwartz C., Fontes M.;
 RT "Determination of the genomic structure of the XNP/ATRX gene encoding

RT a potential zinc finger helicase.";
 RN Genomics 43:149-155(1997).
 [13]
 RP SEQUENCE OF 860-2492 FROM N.A.
 RX MEDLINE-95179111; PubMed-7874112;
 RA Staylor C.L., Dabovic B., Gulliano M., Gecz J., Broccoli V.,
 RA Giovannazzi S., Bossojaco M., Monaco L., Rastan S., Boncinelli E.,
 RA Bianchi M.E., Consalez G.G.;
 RT "Cloning and characterization of a new human Xq13 gene, encoding a
 RT putative helicase.";
 RL Hum. Mol. Genet. 3:1957-1964(1994).
 [14]
 RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE-94214473; PubMed-8162050; Villard L., Staylor C.L.,
 RA Gecz J., Pollard H., Consalez G.,
 RA Millaesau P., Khrestchatsky M., Fontes M.;
 RT "Cloning and expression of the murine homologue of a putative human
 RT X-linked nuclear protein gene closely linked to PKAI in Xq13.3.";
 RL Hum. Mol. Genet. 3:39-44(1994).
 [15]
 RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
 RX MEDLINE-95211835; PubMed-7697714;
 RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
 RT "Mutations in a putative global transcriptional regulator cause X-
 RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
 RL Cell 80:837-845(1995).
 [16]
 RP SEQUENCE OF 1375-2492 FROM N.A.
 RA Pearce A., Chapman J.;
 RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.
 [17]
 RP E2H2 BINDING.
 RX MEDLINE-98167853; PubMed-9499421;
 RA Cardoso C., Timsit S., Villard L., Khrestchatsky M., Fontes M.,
 RA Colleaux L.;
 RT "Specific interaction between the XNP/ATR-X gene product and the SET
 RT domain of the human E2H2 protein.";
 RL Hum. Mol. Genet. 7:679-684(1998).
 [18]
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
 RP HETEROCHROMATIN.
 RX MEDLINE-20040663; PubMed-10570185;
 RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
 RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
 RT "Localization of a putative transcriptional regulator (ATRX) at
 RT pericentromeric heterochromatin and the short arms of acrocentric
 RT chromosomes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13986(1999).
 [19]
 RP DISEASE.
 RX MEDLINE-20213147; PubMed-10751095;
 RA Villard L., Fontes M., Ades L.C., Gecz J.;
 RT "Identification of a mutation in the XNP/ATR-X gene in a family
 RT reported as Smith-Fineman-Wyers syndrome.";
 RL Am. J. Med. Genet. 91:83-85(2000).
 [110]
 RP VARIANT ATR-X SER-1713.
 RX MEDLINE-97196774; PubMed-9043863;
 RA Villard L., Lacombe D., Fontes M.;
 RT "A point mutation in the XNP gene, associated with an ATR-X phenotype
 RT without alpha-thalassemia.";
 RL Eur. J. Hum. Genet. 4:316-320(1996).
 [111]
 RP VARIANT JM GLN-2131.
 RX MEDLINE-96224392; PubMed-8630485;
 RA Villard L., Gecz J., Mattel J.-F., Fontes M., Saugier-Verber P.,
 RA Munnich A., Lyonnet S.;
 RT "XNP mutation in a large family with Juberg-Marsidi syndrome.";
 RL Nat. Genet. 12:359-360(1996).
 [112]
 RP VARIANTS ATR-X.
 RX MEDLINE-97467722; PubMed-9326931;

RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Azenbauer B.,
 RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,
 RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,
 RA Higgs D.R.;
 RT "Mutations in transcriptional regulator ATRX establish the functional
 RT significance of a PHD-like domain.";
 RL Nat. Genet. 17:146-148(1997).
 [113]
 RP VARIANT ATR-X LEU-246.
 RX MEDLINE-20123062; PubMed-10660327;
 RA Fichera M., Romano C., Castiglia L., Falla P., Ruberto C., Amata S.,
 RA Greco D., Cardoso C., Fontes M., Ragusa A.;
 RT "New mutations in XNP/ATR-X gene: a further contribution to
 RT genotype/phenotype relationship in ATR/X syndrome.";
 RL Hum. Mutat. 12:214-214(1998).
 [114]
 RP VARIANT SHS LYS-1742.
 RX MEDLINE-99347960; PubMed-10417298;
 RA Lissi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,
 RA Prieto F., Fontes M., Martinez F.;
 RT "Mutation of the XNP/ATR-X gene in a family with severe mental
 RT retardation, spastic paraplegia and skewed pattern of X inactivation:
 RT demonstration that the mutation is involved in the inactivation
 RT bias.";
 RL Am. J. Hum. Genet. 65:558-562(1999).
 [115]
 RP VARIANT CWS THR-2050.
 RX MEDLINE-99326061; PubMed-10398237;
 RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,
 RA Curtis M.,
 RT "Carpenter-Waziri syndrome results from a mutation in XNP.";
 RL Am. J. Med. Genet. 85:249-251(1999).
 [116]
 RP VARIANTS ATR-X E-175; 178-V--K-198 DEL; S-190; P-219; L-246 AND C-249.
 RX MEDLINE-99219535; PubMed-10204841;
 RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Beloune J.,
 RA Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;
 RT "Evaluation of a mutation screening strategy for sporadic cases of
 RT ATR-X syndrome.";
 RL J. Med. Genet. 36:183-186(1999).
 [117]
 RP VARIANTS ATR-X S-179; L-190; I-194; C-246; F-1552; S-1645 AND C-1847.
 RX MEDLINE-20451413; PubMed-10995512;
 RA Wada T., Kubota T., Fukushima Y., Satoh S.;
 RT "Molecular genetic study of Japanese patients with X-linked
 RT alpha-thalassemia/mental retardation syndrome (ATR-X).";
 RL Am. J. Med. Genet. 94:242-248(2000).
 [118]
 RP FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
 RP GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN
 RP BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.
 RP SUBUNIT: PROBABLY BINDS E2H2. BINDS ANNEKIN V IN A CALCIUM AND
 RP PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSERINE-DEPENDENT MANNER (BY
 RP similarity).
 RP SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
 RP HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
 RP INTERACTING WITH HP1.
 RP ALTERNATIVE PRODUCTS: 5 ISOFORMS: 1, 2, 3, 4 (SHOWN HERE) AND 5;
 RP ARE PRODUCED BY ALTERNATIVE SPLICING.
 RP TISSUE SPECIFICITY: URICIOUS.
 RP DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-
 RP THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X
 RP SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPRISING SEVERE
 RP PSYCHOMOTOR RETARDATION, FACIAL DYSMORPHISM, UROGENITAL
 RP ABNORMALITIES, AND ALPHA-THALASSEMIA. AN ESSENTIAL PHENOTYPIC
 RP TRAIT ARE HEMOGLOBIN H ERYTHROCYTE INCLUSIONS.
 RP DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF CARPENTER-WAZIRI
 RP SYNDROME (CWS), AN X-LINKED RECESSIVE CONDITION CHARACTERIZED BY
 RP MODERATE MENTAL RETARDATION, SHORT STATURE, BRACHYDACTYLY WITH
 RP EXCESSIVE SKIN CREASES, AND WIDENING OF THE KNUCKLES.
 RP DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERG-MARSIDI SYNDROME
 RP (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY
 RP SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS,

[illegible]

Db	536	M-----	EWGSSVDDHGGDSSSGTEGEVESSVKLNISSKDNRRGIGKSTTAKVTKELY	588
OY	1249	K-----	TKKAAEKTECKTYNDLIKENDKKOYEDCHKPKNSNGYDWOCCGINLVED--	1302
Db	589	KLTVPISLPNSPIKADQO---	EWPODKDYGKSC-GINPKLEKGGIQQEJNSDEHJLVEYEV	644
OY	1303	-----	PRVCMPRR-----OKLCYHLAND-----	1322
Db	645	SLLLEESDRLRSPRVKTTPLRLPRPTENPVTNSNDECECNTEVKEKQLSTVPYAKKRKRRSS	704	
OY	1333	-----	NEIKKLOQOVNIKE-----AFIKSAAEFFSWYKYSKDGEJNEL--	1363
Db	705	DSADINPKPNKLPKSKQSEITYDQNSDDEMLATILKGVSRKSHSS-----	SDPTDINEIHT	759
OY	1364	-----	DKELKEGIIPIPAFLRSMETTPGDRDLFGTDLISKGHEG-----SKLKQIDSL	1413
Db	760	NHKTLYDLTKTOAGDKDKGKKRRKSSVSGS--DE---	DKKGSAKSIIISKRRKQOSE	813
OY	1414	FKNSDQSPNKGTRQEWMTESHEIMEAMLCALYKIGAKR-----	DDF-----	1456
Db	814	SSNDSE-----	LEKELKSMKIGARTTKRRLPNKDDPDSSEDEKH	855
OY	1457	-----	TENGYNNVKS-----DKSTLEEFARK--POFLR	1485
Db	856	SKKGADNOGHNLNLTISOEGSSDDEARERKQERTFPSSAEGTVDDTITIMEIREDLPRKQOAS	915	
OY	1486	WLTEWYDYCYTRKY-----	LKQVOEKCSNDOLKCDTECNKCEBYV-----KYMKKKK	1536
Db	916	ASTGVDKLSKEDSFTLSLEYKVAETEKESKILK--ITCKKVDGSLDIAEKLKADQO	973	
OY	1537	EWIPDQKYYKDERKRFDRQHIGVAVYDTGTNAIDYLNRRFTASCGDKPGSASAVORN	1596	
Db	974	-----	SDTESED--DKQO-----SKNGTE-----	990
OY	1597	IQLLEKAVYDADHGCCTFEJENDDKYTNISKDKCGLYEAMGAIKMQNKPNNY	1656	
Db	991	-----	EKKPSPDFK-----KYIMBOQYESS-----DGTE	1017
OY	1657	NIKELTEDVLEPSPRHLRICHALDGNVTPEVDENGELKRLM-----	EVAATEGYNL--	1709
Db	1018	KLPREHEICHHPKGIQIKNGTJDGEKKSKINDITSKKDELSDYAEKSTGDCSCSS	1077	
OY	1710	-----	GQYKKEK-----EKERTJSDAKHYSYEVPPGSAMKYSTYDLRDIIL	1755
Db	1078	EDKSKNGAAGREKRCRKLKLGKSSRRKQDCSSDTEKYEMKGDGNCSS-----	DKR--LK	1130
OY	1753	GIDNLEBKQTEENLKIKFNKNGTSVGKSDSTEN-----	PGSTARFPMNKECY	1806
Db	1131	RIELPERNLSSKNTYET--QSGSSSSDAEESSENKKKKQRTSSKKAIVYAKKR--	1186	
OY	1807	WNAMICGYKRRDNGNSGNSAR--SDEDLKKGSPSDDD--	YPMGNRDEGTAOFLWF	1863
Db	1187	NSLRTSTFKRQADITSSSSSDEDDDNIGSGSSDEQKIRPVTE-----	LVLS	1236
OY	1864	AEMGEDCCKKH--EKELEKLVGAENDYTQGNENKRRKCTDQYKFFISMKPOYEQI	1922	
Db	1237	SHTG--FCQSGSDALSKSPYVYVDDDDNDENRHA-----	KMLLE-----	1278
OY	1923	KRYGENDKTIYSEHPYAKADEAREYLDKOLKTKICNKGDEYKCMKDVSTQRLTDGNS	1982	
Db	1279	-----	EIKANLSSDEBSSD--DEPEBGKKTGQNEENPGEDEAK-----NQ	1319
OY	1993	QNMASLDDPEKEVEGECNCQVRRGPRVARRPTSPRSVLISKATASK-----	KEAKTAP	2038
Db	1330	VNSSDSDSE-----ESK-----	KPRYKHRL-----LRHKLTVSOGSEGEKKTRP-	1360
OY	2039	TJQEKYKENTLITTEKRAQTFRRAAQOTRKRSTATTATESDV--GIWYKALISNPKPSRGI	2097	
Db	1361	KEHEKVG-----	KNRRKVSSEDEBDDPOESGVSEYVSSBEO-----	1400
OY	2098	EGCNPRTYGOYPKWGCIVGSKSNEJGICMPRRKRLCTINNIOYLVNTEFNKRRNDIKEA	2157	

```

Db 1401 ---RPRTSARK---AELEENQSRYSKQKKRRIRIKVOE---DSSSENKSNSEEBE 1447
QY 2138 FIKCAIIEFOFIMLYIENPAENELONGTIPDEFKRMITYGYDKDMFGTDISNDK 2217
Db 1448 EEEEEEEEEE-----EEEEEEEDENDSKSPGGRKRI-----RIKLKND 1487
QY 2218 KLIJTVNSTVTLINENKKNKQDKKDELRKIFWEKNNKFIEWGMVIGLYTLHLDENEKE 2277
Db 1488 KIRJETONML-----KEEERKRRIAEI-----ERERE 1515
QY 2278 KIRDNVOYND-----MTKLTPSELFEVRKP-----QELRMFT 2309
Db 1516 KIREVIEIEDASPTKCPITTKIVLDEDETEKPIVQVHNHMYIKLPHROVDGQF----- 1571
QY 2310 EKAEEFCNKRKQOLIKLENGCKEYCDNGSNDGKTOECAE-----ACVTV 2353
Db 1572 -W-DCCCESEVKTKKSPSGCILAHCML--GKTLQVVSFLHTVLLCDKLDEFTALVCP 1627
QY 2354 QNFIKKWTVEYEROREKFKKDDGKKKYDPTERDIEKATCAHEYLANKLKELCGNKDC 2413
Db 1628 LMTALNMNNEFEKWOGLKDE-----KLEVESLATAVK-- 1660
QY 2414 SCMKPSSOLPRTQOOSDANDMPESLDVYPEEFNKCEPELSKSGSIHTKKITEPK 2473
Db 1661 -----RPOERSYMLQRMQ-----DGGVMT----- 1680
QY 2474 IPMNCVEKAAVYLSKEAENNMDITLKEFIPIESTREKESKNSWTNNPCDPKPYAPDK 2533
Db 1681 -----IGYEMYNLAQGRVVKSRKLEIF-----NKALVDP-----GPD- 1714
QY 2534 YIGRRNPENRENREKFKYDIEKCKNSKFOEKKRVCVPRR-----EHM 2579
Db 1715 -----FVVCDESHILKNEASAVSKANSISRRRIILGTPLQNNLIEYH 1759
QY 2580 CLRNDEIKIEILKDSNYLLKAVRTARNEGIDITKNFSENGCAMPICD--TMK----- 2633
Db 1760 CMVNF-----IKEN--LLSlike-FNRFINPIQN-----GQCADSTVWDVRYMKKRAH 1805
QY 2634 YSFADLGDIVRGTDMLRIGGYLP-----VEIKLYKF- EYIYKWRNKNKR 2680
Db 1806 ILYEMLAGCVQRRDYTALTKFLPKHEVYLAVRMTSIOCKLYOYLDHLTGNNSEGR 1865
QY 2681 NK-----YNDVOTFRSAW---WDANKRDIWKAMTKCAPEDAKLFRGRMDGFERITLIOD 2732
Db 1866 GAGAKLFODFQMLSRIFWHPWCLDLYISK-----ENKGYFDEDSMDEF---IASD 1914
QY 2733 KCGHDDPPVDYIIPORFRMTSEMSEYYCKALMELEKFKKSCDHCKTSDRCKND----- 2787
Db 1915 SDETSMSSSDYTKK-----KKCKKGRK--DSSSGSGSDNDVYEVIK 1956
QY 2788 -----YDEN-----KCEQCKTRCOE-----YKNFVLKMKSLFDIQ 2817
Db 1957 VNNSRSGGEGENVDETGNPNPSVLKLEBSKATSSSPAPDMYKDFVTADAEVLEH 2016
QY 2818 SNKYKEIYE---OPITYIKISTY-----DHVQNFVOKLTKFKSECVSESEYELIET 2865
Db 2017 SKRMVLEIEILMAEIEIGKVLVFSOSLISLDLIEDFELASREKTE-----DK 2065
QY 2866 SKCLNYKFNENDGSSNIRTYAFEEETPKSYKEACSTLPKSNPLDNCPTDONKDGCKELQ 2925
Db 2066 DKPLIYK-GECKWMLNIDYR-----LDGSTYASRK--KMAEE 2101
QY 2926 FTFCSSKNDYD-----NLDNMNAYLVL-----NSSDDNKGVILIPRRRH 2964
Db 2102 F-----NDETNRGRBLFIISTAGSGLNLYVANRVIIFDASMNPSYDIOSIFRYVRFQ 2156
QY 2965 LCTRPITANV-----RKDGKELIKKLLTSAFSGOLLGQKKSEBELCFEAMKYS 3016
Db 2157 -TKPYVYRFLAOGTMDKIDYDROVTKOSLSFRVYDQOV--ERHFTMNL--TELTY 2209
QY 3017 YADYSIDILKTMDSLSSEKIKIFETSEATENKRTWMENNRQIMHAMLCGYKIATS 3076
Db 2210 FE-----PDLDDPNSEKKK-----RDTFPLMKOYIILAEILOIHKEHIVGHEHDS 2256

```

```

QY 3077 KYTLDEGMCOLKRDERTNOFLNMLIEMAKOACKEKHVSDSLTKCPKRSNEDNFEASELL 3136
Db 2257 --LLDH-----KEEBELTEERKAMAEYEA-EKKVL--TMRFINPTGTN----- 2296
QY 3137 RQPGCONDIRKTYISLNLIKNTMENLNIKYKOLKDOSSGNIDNKPSEENVQSYIKSKSQ 3196
Db 2297 LPPVSFNSQTPYIIPNGLGALSMSN-----QOLED-----LINQGREK---VEATNSV 2342
QY 3197 CALELDINEIVTGT--KNEN 3216
Db 2343 TAVRIQPLEDIISAVWKENMN 2363

```

Search completed: April 28, 2003, 10:30:09
Job time : 116.033 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 65.5138 Seconds
(without alignments)
5197.501 Million cell updates/sec

Title: us-10-087-013-2

Perfect score: 19407
Sequence: 1 MGFSCRYFLIKMGNAASSLE.....IDLINDSLVJLNLILYMKY 3542

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
al number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	4515.5	23.3	2706	2 T28155	variant-specific s
2	4254	21.9	3006	2 T28625	variant-specific s
3	4010	20.7	3026	2 T28431	variant surface pr
4	3489	18.0	3078	2 T28432	variant-specific s
5	3130	16.1	2924	2 T18378	variant-specific s
6	2714.5	14.0	2664	2 T28626	variant-specific s
7	2686.5	13.8	2647	2 T28161	hypothetical prote
8	2677	13.8	2212	2 T28157	erythrocyte membra
9	2303	11.9	2135	2 T14602	variant-specific s
10	1749	9.0	2197	2 B71600	variant-specific s
11	1743	9.0	2182	2 T28634	variant-specific s
12	1689.5	8.7	2228	2 T14029	variant-specific s
13	1574	8.1	1711	2 C71625	variant-specific s
14	1523.5	7.9	2042	2 T18396	variant-specific s
15	1249	6.4	1729	2 T18396	erythrocyte membra
16	811.5	4.2	431	2 T17600	variant-specific s
17	617	3.2	1435	2 A37793	erythrocyte-bindin
18	548.5	2.8	4550	2 T18440	hypothetical prote
19	523.5	2.7	440	2 E71625	variant-specific s
20	520	2.7	3394	2 T18501	hypothetical prote
21	491.5	2.5	3724	2 T18427	hypothetical prote
22	435	2.2	2829	2 A42771	reticulocyte-bind
23	433	2.2	2533	2 T18477	hypothetical prote
24	423.5	2.2	4981	2 T18489	hypothetical prote
25	421.5	2.2	1979	2 C71622	hypothetical prote
26	420.5	2.2	2401	2 T28676	hypothetical prote
27	415	2.1	6669	2 S55024	nebulin, skeletal
28	405.5	2.1	2269	2 T28677	thorphy protein -
29	404.5	2.1	2510	2 T28160	hypothetical prote

30	403.5	2.1	2657	2 T18497	hypothetical prote
31	384.5	2.0	2573	2 D71614	hypothetical prote
32	381	2.0	3135	2 A45884	transmission block
33	378	1.9	248	2 C71624	variant-specific s
34	368.5	1.9	2485	1 H71621	serine/threonine-s
35	362	1.9	1819	2 A71928	cag island protein
36	358	1.8	1939	2 T18372	repeat organellar
37	355.5	1.8	2295	2 B71621	probable membrane
38	354.5	1.8	1045	2 T18373	erythrocyte bindin
39	353	1.8	3973	2 B71612	hypothetical prote
40	352.5	1.8	1927	2 G64585	cag pathogenicity
41	347.5	1.8	3844	2 T18402	asparagine/asparta
42	347	1.8	4688	2 F82885	hypothetical prote
43	344.5	1.8	1526	2 A45605	mature-parasite-in
44	342.5	1.8	1070	2 T30848	Duffy receptor - p
45	338	1.7	2500	2 G71609	hypothetical prote

ALIGNMENTS

RESULT 1									
T28155									
variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum) (fragme									
N:Alternate names: erythrocyte membrane binding protein 1 (EMPI)									
C:Species: Plasmodium falciparum									
C>Date: 15-Oct-1999 #sequence_rev15on 15-Oct-1999 #text_change 20-Jun-2000									
C:Accession: T28155									
R:Rowe, J.A.; Moulds, J.M.; Newbold, C.I.; Miller, L.H.									
Nature 388, 292-295, 1997									
A:Title: Plasmodium falciparum rosetting is mediated by pFEMP1 and requires complemen									
A:Reference number: Z20477; MUID:97373957; PMID:9230440									
A:Accession: T28155									
A>Status: preliminary; translated from GB/EMBL/DBJ									
A:Molecule type: DNA									
A:Residues: 1-2706 <ROW>									
A:Cross-references: EMBL:Y13402; PIDN:CAA73831.1									
A:Experimental source: strain IT 4/25/5									
C:Genetics:									
A:Introns: 2493/3									
A>Note: R29R+var1									
Query Match									
Best Local Similarity 23.3%; Score 4515.5; DB 2; Length 2706;									
Matches 1191; Conservative 372; Mismatches 871; Indels 1315; Gaps 123;									
OY	34	SARVLEERYAKNIRHPK	---YAKEHYDSLKGDITKEF	-----RGSPSTPVVK	79				
DB	14	SATDVLEKIANGIYNCKE	KVYPE--NELKGLISNAIFVDOLRKEINTISPGSD	----	67				
OY	80	HNYYTPPCNDLHKEHTN	LRYDVLNLRHPCHGREONRFEDESEECG	---NKIRYKRNDA	138				
DB	68	-----SCSLDHKPHN	INTEYTEGRKPCYERNEKRFSGNEGAKCSGSDIR	DIYGISAG	120				
OY	139	IACAPPRRMHCDKNLEAL	DINTONITHDLGNLVYAKTEGSGIVNNHHPKGT	----SD	194				
DB	121	GACAPFERROMCDNRNLE	TYLNTKNTNTHTDIGNLVYAKTEGSGIVNNHHPKSSG	KRSS	180				
OY	195	ACTALARSFADIGDIV	RGIDMFKNPVADKYETGIGREYFKIHQMEDEVANDY	---NPDSG	253				
DB	181	ICTALARSFADIGDIV	RGIDMFKNPVADKYETGIGREYFKIHQMEDEVANDY	---NPDSG	240				
OY	254	NYTKLRAMNVRNKNKY	WEAITCDASYKSGYFMQSESNPLF	---SNPKCGHKGQVPTNLD	312				
DB	241	NYTKLRDMWMAINRREY	KAITCRAPMEANFERNISGNMKAFTSGYCGHSE	TNPVNTLD	300				
OY	313	YVPOYLWFPDMEG	EEFRKRNKIKKKYKDCGRNDKELTCSHNCDDTTITW	KKGIHL	LD	372			
DB	301	YVPOFLWFPDMEG	EEFRKRNKIKKKYKDCGRNDKELTCSHNCDDTTITW	KKGIHL	LD	360			
OY	373	NKCDCTSKCFVFEW	MGNOEAFKROKERYEKESYLSLNDNFVNI	---NSEYKQFE	431				
DB	361	LNCPRCENACSN	TKWIEIQRKQPFDRKRRKTMET	-----KIKTNISNNEKDEFEYE	412				

QY 432 KIKETQVATNDTFLNLLNEGYYCKGGLGEKDIPTFNSADKGIPIYSEYQVCPDQGVK 491
 Db 413 NLDKGYSTINTWFLSJLHNGOCODNIDKKNKTFKNLE---TFGPGYCEACPITIGVK 469
 QY 492 CDGIKYTHSONDRVRVNNEDYKPPW-----GAKPTNTIVYSGNEQDITOKLEFNCS 546
 Db 470 CSNEKCTVTEHEW-----NSNNRLPTDTSTKLNKATNIDMLVNDGIGALIDNELEKNOTK 525
 QY 547 STNYKDNKNQKWCYKXKDNENIRKLEONTEIN--DNPKITSEHNFELWVYLLRDT 603
 Db 526 YGILKIGIKKQKQCYLN-NIDOCKI--NNVANSGYFDNK--IAFNVLFOHMLRYFYBDH 580
 QY 604 IKMNDKLTCT---NNTTHHCIDECNRKCTCPDRWVYKKEEEMNSIKKLTFTKKNKIQ--- 657
 Db 581 NMLKERRIDVCKEKNINENICIKREKTCCEGVKLEKKEEMDKLNOCHYOKNKHIMFL 640
 QY 658 -----QSYYSNINLFEGEFFKVMKDLKDEAKKELMENIKRKNERSNLENNRDYLEN 712
 Db 641 IYWTITGEFEKIT--FPNDFKALDDVDIIN----- 669
 QY 713 AIELLDHLKETATYCKDNNTNEACETSHNATTPNCVPRGSTOPTKNIKELAOYFKRSA 772
 Db 670 -----VLDITLKE-----CODTH----- 681
 QY 773 YEBARRGJHLKGAHNEGITYKRGRRKDFKDNLCRIMI KHSNRMLGFSNGPCKGTGD 832
 Db 682 -----CKI----- 684
 QY 833 GIOTREVGTEWEVDPEHNRKDEHYIMPRRRHICTSNLEHLQTDHPLNGNIYDLVN 892
 Db 685 -----EKIRSIDVDLI----- 695
 QY 893 NSFLGYLLSAYEAKKIIRMKERKNLKGPEYNDPKHQTICAIRYSFADIGIING 952
 Db 696 -----KEIIS----- 700
 QY 953 ROLMERNGDKVQLQGLETFVGNHSLKGNKYNDAFYLLKRENMWEANRAKVE 1012
 Db 701 -----WLONKLEV--- 708
 QY 1013 AMKCDIRKYLKDSGHOSSTGYCYSDHPPDDYIPQKLRMTAEAKCYQAKREYDL 1072
 Db 709 ---C-----KSHHEDKHEYC-----CDILPKSVDDD 732
 QY 1073 KECKEBCDKDNGGCTKESGCTCTCTEACNEYNIDILMKQMNILSDYKELHEQAO 1132
 Db 733 EEDDEEV-----DEEKE----- 744
 QY 1133 MSVNSGIEASSTAKNHIDRNVLIEFLSELQONGSKSNSGTSDESANVIGTNTTYENGA 1192
 Db 745 -----ESSQTTKRN-----SOKGTFKASGVGACAIIVGV-L 777
 QY 1193 YLHDYGNFDDCOSONECCDEKSDGKONEKAFKPDQDDHAGAGCKSGSPFTVQIKTK 1252
 Db 778 QOKSNGSIDCNKKNR----- 793
 QY 1253 KAEBKDECTVNDILKENDGKROVEDCHPKKNSNGYPMQCNINLVE--DPRVCMPPRR 1311
 Db 794 -----KKN-----EMGCDKNFTVDNEGVCMPRR 818
 QY 1312 OKLCVHEFLANDNEIKLOSQVNLKEAFIKSAAETFSWYIYSKDSGEGHLDKELEKGR 1371
 Db 819 KSICIHMLTLEBQTK--NKYQLEAFIKCAAKETNLMKRYKKNKDEASEL--LKKGK 872
 QY 1372 IPPAPFLSMFTTBDYDFLEFGTDISKGHGBSKLKEQIDSLFNGNOKSPNGKTRQEM 1431
 Db 873 IPDEFMRIMFTTDFDDEFLCENDMGK--DVDAVKKNINIKVENNSSKRGFKKATDPPNMW 929
 QY 1432 TEHSHEIWEALCALVYKIGAK-----KDDFTENYGYNNVYFSDK-----STLLEFAKRP 1481
 Db 930 NENGPOJINGMICALIHADYDOSTIKND--NYKIEKVTILAKRDSNGMTLSEFAKPP 985

QY 1482 QFLRWLIEWYDYCYTROKYLKDVQCKRSND--OLKCDTECNKCEYKYIMAKKK-EM 1538
 Db 986 KFLRWLEWWDYDCKRQKYLTEVASTCS IDQGLKCDRCGNKCKDEYKMKRRKKEEM 1045
 QY 1539 IPDDKIYKDEBKKRDRQHIGVAVDYGTNATDVLNKKFPASCCDRKGS-----SVY 1593
 Db 1046 NLDKRYKDRKREKNGIDKGIIVADYLANKEYLKFKFPAKSCVTSKGKONSATEYV 1105
 QY 1594 ORNOLLEKQAYDADKHCQCTFIENDKYTNISCKDKGLVKEBANTGAIKWONKGN 1653
 Db 1106 KKNLELSEGOYIDAQYCCCTFI--HDKYKSTISGRNCCGLNSAPKAKNIKWRNSDR 1164
 QY 1654 NYNNLKE--LTDVLEPSRLRICFHADGNTYDPEYKDENGIKRKLMEVATEGNLQ 1711
 Db 1165 DYAFKRRNLISGVDFPFSRRLRICFHADGNTYDPEYKDENGIKRKLMEVATEGNLQ 1224
 QY 1712 YKKEKKEKKEIKTSDAKYEVPCGAMKYSFYDLIDILGIDNLEDEKOKTEEMIKKI 1771
 Db 1225 YKKEKKEKETEAEAKYIETEVQPCAMKYSFYDLIDILGIDNLEDEKOKTEEMIKKI 1284
 QY 1772 FNNKGTSGKGSSTGTGPNGSTARKFPMENKRCVNAAMTIGYKRGHDGNSGNSARSDE 1831
 Db 1285 F-----KSESNESQGRQTFWNNNKGCVWEMAKCGYKHGRDGNSSANSASDQ 1332
 QY 1832 DLKCGSYPSDDYIPAKKNDEGATAYOFLRPMFAMGDEPCKHKEKELKLVGACNDYTCG 1891
 Db 1333 DLKRCDSVPSDDYIPKGRDEGATAYOFLRPMFAMGDEPCKHKEKELKLVGACNDYTCG 1392
 QY 1892 DNEBKRRKCTDCTQYKRFISEMPOYKQIKKYGKNDKIYSEHPPAKDAEDAREYLDK 1951
 Db 1393 DNEBKRRKCTDCTQYKRFISEMPOYKQIKKYGKNDKIYSEHPPAKDAEDAREYLDK 1452
 QY 1952 QLKRICENKSGDCEYKCKKDVSTORLTDGNSQNNPASLDEPEKVEKCNQVPRGPRV 2011
 Db 1453 QLOKSC--NSGKCD--CWNKKT--SNGNMPASLDEPSTYKDCBQCPPPP-- 1502
 QY 2012 RRETPSPRVSILSKATSKKEAKT-----APPTKOPK-- 2044
 Db 1503 ---PPAPRPPAPRSGNDYGRSEBPGDGLPLPLPPPPPKGCGGAGRILSTPRNG 1559
 QY 2045 --VENILTEMBRQOTRRAAOOTRRKRTSTATTESDVQWNAILSKPDSRGCI--BGN 2101
 Db 1560 TIEDEEDDDEGEKAAEAASETESK--EVEKEDETTEKCKIVANILTKGMIDGACN 1617
 QY 2102 PRTYGOYRK--WCIY-----GKSENENGJCMPPRRKIKC 2135
 Db 1618 OK-YG-YPORHGMKCVPTTSSSTSERGASRNKRNLDSTKSGKSGITGIPRRKRLY 1675
 QY 2136 INNIO--YINYETE-----NKRDNK-----EAFKCAAIETOFPLM 2171
 Db 1676 IKKIDMASGNTQAGSNTGSGDSTGASSNPQONGVSTSPYALLHAFAVESAAVEFLMD 1735
 QY 2172 KY-----ITENPAE-----NELONGITPDEFRIMYYTGYDKWDFEG 2210
 Db 1736 RYKKEKEIEKQOQOETGLVASETSEDTEHPQKLOSGTIPDLFQMKOYTLGDRDLCVG 1795
 QY 2211 TDISNDKKITTYNSVTTLNENKKNOD--KKRD-----BELRKIPWKKKKFTWEGMI 2263
 Db 1796 ---KTPDGDITVSAKDQIMDKIOAKIOILPKKTPPSVYKTPPOOTWNNKKAESITWNNMI 1852
 QY 2264 YGLTYHLTDENKERR-----DNTQYNDMYL----- 2291
 Db 1853 YALTYK--TTPSGEKRKQJPEVTKLFLDEKGTPOGNSKRYQYKVY-KLEEBETSGAPKSTE 1910
 QY 2292 --TPS-----LEEFYKRPQFLWFTBMAFEPCNKKRQKILKLEAC-----KEYCNG-- 2337
 Db 1911 SSSPSGENTPLNDFTSRPPYFYLEMGONFCKEKRRKRGIDKISNCLDGDQYSGDSEA 1970
 QY 2338 -SN-----DGTOCAEACVYQONFIKKWTEYEROREKPKKDKGKKYK---DY 2383
 Db 1971 CSNIDVNDKIFADLEGPYKAPCSSYRKWITGKTYQYKOKRAEBOQEGCKKFTNNHY 2030
 QY 2384 PSTERDIERATCAHEYLNNKLKELGKNKDCSCMQKPSQLPPTTQOOSOSDANDMPESLD 2443

QY 661 YSNINLPECFYKVDKLDKDE-AWKEIMENIKRKKNEFSLNENRDLLENALIELLD 719
 Db 648 YNNLSYFDSFLQYAFALDDEKGMDFTEDLK-KKEPESKTPNPTGKSOAIEFLD 706
 QY 720 HKEFATICKDNNTNACETSHNATNTPCVPKPGQOPPKNIELOIYQKRSAYEABARR 779
 Db 707 HIKDMALTCRDNNNSNFCVSKKVTNPGCKNPSASNNLVRKRLAEMORARKEARR 766
 QY 780 GLH-KIKGAHEGIRYGRRRRDKDNLCRIMIKHS--NRNLGS-NGPCDGK-GTGDG 833
 Db 767 GGEINLKGDATKGTYROGPADGEK-NVCSINQNHNTVQNNNRAIYQGPCGKOSNG 825
 QY 834 IOTRFVVGEMEYDEPHMKRDEHDVIMPRRRKICISNLEHLOTDDHPL-NGNIYDDLVA 892
 Db 826 V--RMKIGTPWK-PGRQIQMSAEDIYMPRRQHMCTSNLEYLOTGKPGPKYGD--GLYV 880
 QY 893 NSFLIGDVLISAKYEANKIIRMYEKENNLGPKREVTPDKQJOTICRAIRYSPDIGITIG 952
 Db 881 NSFLIGDVLISAKMDAKIIEIYKKONN--KSLNIDPEENESACRALRISFADLIGITIG 937
 QY 953 ROLWERNMGVAKLOGLHLETFVGNHRSKLGKNDKYNDDAPRYLKLRENMWEMAKRAVME 1012
 Db 938 ROLWERNMGVAKLOGLHLETFVGNHRSKLGKNDKYNDDAPRYLKLRENMWEMAKRAVME 1012
 QY 1013 AMKCDIKYLDKSGHQSIOSSYCGYSDHTRPLDITPQIRMTTWAEMTCAYQKKEEYDL 1072
 Db 998 AMKCDIKYLDKSGHQSIOSSYCGYSDHTRPLDITPQIRMTTWAEMTCAYQKKEEYDL 1072
 QY 1073 KKKCKCKDNDGOGCTKESGTGCTKCTACNEEYNDIIGLAKRONNITSDKKELEHGO 1132
 Db 1046 VADCKSCCK--AKSCIOGD--DCTKCKAAGNMYKKIKIPWEGOEKIKKNAOILKRL 1102
 QY 1133 MSVSGIEASTANKHIDRANIEFLSELYOON-GKSNKSGTSPESAVITGNTTYENVG 1191
 Db 1103 DSVNKESSKKTADADAOYVHFLAELIRKSGGKGNVKTVPPTTPTNTLYSSAA 1162
 QY 1192 AVL-HDTGNFDGOSQNEFCEKSDGKDNKRYAFRDKPODHGAGGCKS-GSKPRVQIK 1249
 Db 1163 GYIHHELRTVGCNTQKCEYCSK--NGKTAEPDKRYEAKCNDRNP--Q 1212
 QY 1250 TKKKAEEKDTECKIVNDILKENDGKQVEDCHP--KKNNGYPDWQGINILVEDPRYCM 1307
 Db 1213 PAPKEDDEDA--CDVYKPLKDKGETDDIDGCKQKAGKDKYPMGDCNSQIHTHNGACM 1271
 QY 1308 PPRROKICLVHFLANDNEIKKIOSQVNLKEAFIKSAAEFTFSWYTKSGDEGNELEKEL 1367
 Db 1272 PPRROKICLVHFLANDNEIKKIOSQVNLKEAFIKSAAEFTFSWYTKSGDEGNELEKEL 1367
 QY 1368 KKKIPPAFLSMETFGDYDFLEGTDISKGHGSKLKDIDSLF-KNGDOK-----S 1421
 Db 1326 KNGNIPBEGFKOMYTFEDYDIFFGDISF-HAIVSGVSKVITILEKENDAKYAAKON 1384
 QY 1422 PNGKTRQWMTHEHETWEAMCALV-KIG--AKKDEFTENYGNVNFSDK-STTLBEF 1477
 Db 1385 SNNELLDMMOHOHGDIMEGICALTHKISDEBKKKEIKNKYSYKILNESPGRSKVDF 1444
 QY 1478 AKRPOELFMLEWYDDCYTRQKILQOYK--KSD--QKCTECKKCEDEYKY 1531
 Db 1445 AKKPOELFMLEWYDDCYTRQKILQOYK--KSD--QKCTECKKCEDEYKY 1531
 QY 1532 MKKKREWIPODKYKDEBKKRFRONHIGVWVTDYTG--TNAVYLNKKFLASGCD 1585
 Db 1505 ITKKKV-----EYTK--QKGFDAEKI--TDEKEGEGSTDAISYLAKK--C-- 1546
 QY 1586 KPGASAVQRIQILLEKQAYADAKHGGCTKFTENDDKYTNISSKDKCKGLVEANTGAI 1645
 Db 1547 -----LDDTCMCMQKXVKNTEYVNT----- 1566
 QY 1646 KWQKGPNNYNNLKELEFVLPSSRLRICPHALDNGVTEDEVANGELRLMEVATE 1705
 Db 1567 --PKYTYTN-SNLEKRCB--CQPPQ-----EPPGPEGGAR-----SDS 1600

QY 1706 GYNLQGYKKEKKEKIKTSDAHKYSYEVPCSAKMSYFYDRLIITIGDNLDEKOKTE 1765
 Db 1601 G-----PRTTTPRAGSDARSNVPSPPRPAQDIYHEVAEY-----QEEDEDE 1645
 QY 1766 ENUKITFNKNGSVGKSDSTTGNDPSTARKEFMENKECVNMAMICGYKRGDRDNGSN 1825
 Db 1646 -----DGLDPE 1651
 QY 1826 SARSDLEKKGSDYDDDPYPMGNRDEGTAYOFLRWFAW--GEDPCKHKELEKLYGA 1884
 Db 1652 D--QDEDEYVAGAEED-----LDVGA-RILGRTNSPDEDEDEASEE----- 1694
 QY 1885 CNDYTCGNEDEKRRKCTACTOYKKEFISEWKPOYEKOIYKGENDKIYSEHPVAKD 1944
 Db 1695 -----DDDDDAQOTTEVTGQ-----GEETREHOD 1721
 QY 1945 ABEYDQOLAKICEKSGDEYCKAKDYSTORLTDGNSONMPASLDEPEKEGKNCQV 2004
 Db 1722 TTEETVQO-EKAEEDKDGGE-----TPOKET-----QPKVEVNC-- 1757
 QY 2005 PRGPRVRETPSPRVSLSKATASKKEAKTAPRQPKVYENLTERAOTRTTAAQ 2064
 Db 1758 -----IKTLETTETLEKACPYKYNGRE----- 1782
 QY 2065 TRKRTSTATTESDVTMVKALISNKPDSRGIEGCPNFTYQYPMGCIYKSKENENG 2124
 Db 1783 -----KPPNMWCI--SSGSDASGS 1799
 QY 2125 ICMPPRRKIKNNTOYLYN-----ETENKRDND-IEAKIYKALITOLMLKY-IIE 2176
 Db 1800 ICIPRRKRLHKLIEGVDTVSSDGETTPTDHALKEADIGRAVETPEFLMRYKKIK 1859
 QY 2177 NPAAENLONGT-----IPDEFKIMYTYTGDYKDFG 2210
 Db 1860 EKEKEEELONGHFLPRAQKVSPEDEPHPKKLEKGIPEPRKROMYTIGDYRLDLCVG 1919
 QY 2211 T--DI-----SNDKIIIVYNSVTTLNEN--KKQDKKDELRKIF 2250
 Db 1920 VKDDVAQALIASGDNKSGDNKIKIDISEKIKSVIEKSGQRPGRPGQTTPKPE--W 1975
 QY 2251 WEKNKFEWMEGITGLYHFLD-----ENKKEKIRON-YOYONMTLTS-- 2254
 Db 1976 WOKNGEHNNMICALT-HNIDTROVDQOVAGOLPENGKTPKNSOYQYKAVTTLSSVSN 2034
 QY 2295 -----LEEFVRPOFLWETMAEFCNKRREOLIKLEAG--KEECCGNSNDG-- 2341
 Db 2035 GGPIGNIKLEQFASRPFLRWLEMGEEFCRQYHKLERIKEBHKGONRCDDEGECK 2054
 QY 2342 -----KTOCAEACVYQNFIKKWKTEYEROR--EKKKDKDGK--KYDY 2383
 Db 2095 EMCPCDKSGFETLPCPCAKSGYKRWISRKNDFTKOKAVYEROKRDAEGNNDYKEF 2154
 QY 2384 PSTERDIEKATCAHEYLNMLKELCGNKKDCSMQKPSQJPKTQOQSSANDMPESLD 2443
 Db 2155 SKTLRNTNDA--FLN-SLK--NGPCS-----KNDSDVODEIK 2189
 QY 2444 YVPEE--FNKCEPELSKSGSMITHKTIKTEPIRANCVEKAAVYLSKEANNIDILKE 2500
 Db 2190 FDEERKTFGHETTCPCSK-----IIVKC 2213
 QY 2501 KFIPTESTKKEESKNSMTNNPCDPPKPYAPDKYIGRANCENRENRKRYEWMCKYN 2560
 Db 2214 K-----ENNECDNSP-----NDK----- 2227
 QY 2561 SKFYQEKRCVCPRRHEMCLRLNDEIKIRLDSNYLKMVBRTARNEGIDITIKNENSE 2620
 Db 2228 -----RNINSISAE-----DIEKSNST 2245
 QY 2621 NGCAMPIDC--TMKYSFADLGIWGTMLRIGIYLPVEIKLYKVEFYTKMRNKNR 2678
 Db 2246 QDVTMS-VSDSNTNGKFNFDLNCIKR-----GIFKGI----- 2277
 QY 2679 GRNKYNDVQTFRSAMMDANKDKIMKAMTCAPEDAKLFRKGRMDGFERITL-----IQDKC 2734

Db 2278 -----REDWKC-----GEYGVGDICTLENTNNEERY 2304
QY 2735 GHKDDPPVDYIPQRF--RAMTMESEYCKALMELEKFKKSCDCKSDCKNDYDEN 2791
Db 2305 SAKENDNKKOIIIRVLEFRML-----ESTLEYNKINDKISIC-----MKND----- 2347
QY 2792 KCEOCKRQOEKYNFVLEKMSLEFDIOSNKKYKELYEQPIYTKISTYDHVONFVKLTFS 2851
Db 2348 KKSQCINGCQNCCKEYKMELEKKSEMGKREY-----INQY----- 2385
QY 2852 EGVSEFSEYLETSCMLTKFENNDGSSNIRYAFEEEPKSKYKESCTLPBKNPLDNC 2911
Db 2386 -----RDKNNSN--EAFE--VKSFLETLIPQIP-----V 2409
QY 2912 PFDONK--DGCLEQTFPCSKNDYNNLDWMAYLVLNSDDNKGVLIPRRHLCSTRP 2970
Db 2410 YVDKGHDSITQALKLKCKSEKSEN-----SNEKDYV-----LC----- 2445
QY 2971 TAYNRKGDKEILKKLLTSAFSGQLGQKYKSEELCEAMKYSADYSDIIGKTDMM 3030
Db 2446 -----LLK----- 2449
QY 3031 DTSLSEKIKKIFETSMENETENKRTWENNRRQIWMHMLGCIKATSKYTLDEGWCQLPK 3090
Db 2450 -----LEKAKN-----C-----KD 2459
QY 3091 EETNOFLRWLEWAKOACEKKHVSLSLTKCPNSNEDNEFASSELLRQPCQNDIRKYS 3150
Db 2460 QASGE-----PCQITSENDDDLLE----- 2483
QY 3151 LNLIKNTMENLIKYYKQKQSSNIDNKPSEENVQSYKSSOCALDELINDINEIVTG 3210
Db 2484 -----ENPEAPVNI-----CPKYEPEPVV--EEEKC-----DLAEPAS 2516
QY 3211 TKNNENNEFEVAKLYLPGIYFEDETHKNHVLDCNKEEQYVRKALYFFPHDSFY 3270
Db 2517 ESSTEENSGE-----GSNSEONPSKPEEPPTSETDTP----- 2553
QY 3271 QASLFSTHRAQYD-PKN--DILKSSIVYVLSALGLALHFMKKFKSSVLDRLANT 3326
Db 2554 -PPAPPLIQSQADQPTNSISDLSSTIPGIALALSYFLFKKTKSVLLKRLANT 2612
QY 3327 PCEYGMPTLESKNRYIPYRSGPYKGTYYMEGDTSGDEDKYMDLSSDITSSESEYE 3386
Db 2613 PKGEYMPPLTKSSNRYIPYASDRYKGTYYMEGDS--DSGHYYED--TIDVYSSESEYE 2668
QY 3387 ELDINDIYVPGSPKYYTLFVLEPESKRDLPD--DTPS-----NDPRTIRFLIDEM 3437
Db 2669 ELDINDIYVPGSPKYYTLFVLEPESGNNPTASGKNTPSDTQNDIONDGISSKRTIDNEW 2728
QY 3438 NELKDFVSOYLNPTEPN--NNYKSADIPMNTPEPTLYSDNPEEKPFITISDRDLTYG 3494
Db 2729 NTLKDEFISQYLOSQEPNDVNDYTSNGSNSTNTNITTSNHNVEKRFMSIHDRNLTYG 2788
QY 3495 KEISYININNSTNTNDIPMANRDSYRGIDLINDSL 3530
Db 2789 EBINYNVNM-VNTMDIPIDIRNNVNSGIDLINDAL 2823

RESULT 3
T28431
variant surface protein 1 homolog A4VAR - malaria parasite (Plasmodium falciparum) (frag
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28431
R:Smith, J.D.; Kyes, S.; Craig, A.G.; Fagan, T.; Hudson-Taylor, D.; Miller, L.H.; Baruch
Mol. Biochem. Parasitol. 97, 133-148, 1998
A:Title: Analysis of adhesive domains from the A4VAR Plasmodium falciparum erythrocyte m
A:Reference number: Z20486; MUID:9094502; PMID:9879893
A:Accession: T28431
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA

A:Residues: 1-3026 <SM>
A:Cross-references: EMBL:L42244; NID:g3540144; PID:g3540145; PIDN:AAD0351.1
C:Genetics:
A:Gene: var
A:Introns: 2906/3
Query Match 20.7%; Score 4010; DB 2; Length 3026;
Best local similarity 29.8%; Pred. No. 3, 1e-179;
Matches 1095; Conservative 480; Mismatches 1151; Indels 954; Gaps 127;
QY 30 ESHKSAKNVLEKAKNI-RHPSKYAKEHVDSLKGLDTLAFRCGPSTPVNKHNYIPYPC 88
Db 13 EDDKDAKHVLSIGERYKKEKRYAANYSSQLGTLSNAIFENEPKGOQTEPD-----PC 67
QY 89 NLDKHEHTNLRDYDVLNRHCHGREONRPEDESESGN-KINRYKKNALACAPRRR 147
Db 68 KLIYEHTNVTQKH-GREHCRKRGTERFSDVGGECDDNKKID--SKNNGCACPRL 124
QY 148 HMCCKMLDALDINTONIHDLGNVLTAKYEGESTIVNHPH-----GTSDACTAL 199
Db 125 HLCVRNLNISALDKINNDTLADVCLALHEGQSTIODYPYQAOYASSFSQICTML 184
QY 200 ARSFADIGDVRGIDMKPRVNDK--VEIGLRVYFKIHGMD-EKKNYNDGSGNRY 256
Db 185 ARSFADIGDIIKGRDLYGKKEKLEKYLKFGIYKLDPRAKDHYKDEPDNFF 244
QY 257 KLRKAMNVRNKNVWEALITCDASYKSGYFMQSSNTPLPFNPCKGHCKGK-----VPT 309
Db 245 QLRKEDMNNAROVWALITOHAGESDKYFRKTAAGTGTGOGRCRCDKPPSNDPPT 304
QY 310 NLDPYQYLRWDEMGEBECRKNRIKIKVYKDSR--NDKERLYCSHGHCCTTTWK 365
Db 305 YFDYVQYLRWFEMSEDEFCFKRKHLENAIEKCRGDGSGKBER-YCDFNGFCCKGTAAG 363
QY 366 KGLIHLNCKTDCSTCKVYEWLGNQGEAFKQKKEVEEISYLSNDKKV-----NNI 421
Db 364 KHRYLMDNACAGCFSSCDRKNIAOKDEFEKKNKYDEIR--KNDPTTTTTEGTI 420
QY 422 NSEYKQFYELKEFOYATNDTFLNLNLECKYKCG--GLPBEKDTLPNSADGKIFY 477
Db 421 NNNYRDFYKHLDE-KYKTDAFLNLNLECKEKNHPEVEGKKYIDFNDE--TFS 476
QY 478 RSEYCOVCPDGGYKCDGIKYTHKSDNDRKVN-----EDYKPPWGVKPTNITVYSG 530
Db 477 HTEYREPCPMGIE-----EOKDKWKIRINDHSACKEBELYTPKXNAKYTKINVTSG 529
QY 531 NBOGDITQKLENFNCNSNTNKKDN-----QKWCYCYKDE----- 565
Db 530 EGHEDIKRLKECTCTQONGGSGSDCGNSDSSLCPEWCYQPDLEKYGGGEVDDK 589
QY 566 -----NINCKLEONTIENNDPKIISFNFPELWVLYTLRPTIKYNDKLKTCIN--NTT 618
Db 590 GAGGLCIEFKMEKKEKVK-----KQKTFNPFNFWAHAKSIDMRQTLTCSIDL 643
QY 619 THCIDECNRCLEFDRVYKQKEEWSIKLFTKKKN-IOOSYYSNNINLPEGYFERVMD 677
Db 644 KKECKGKSNCECFKKEKKEKEMIKVQDFNKQDFLEMKHYLYLETILENYEENIQ 703
QY 678 KLDKDEAKKLEMTENIRKKNESNLENNRDYLENALELLDLHLEKATIC-----KDN 752
Db 704 KAYGDLKSIOEMKMKIENONKNRTKD--EDALDVLDEHKEEAEDCLIDHEDDD 759
QY 733 TNBACETSHNATNPVCPKPRGCTOPTNKIEIAOYEFKSAVEBARNR--GLHLKKAHE 790
Db 760 DDCEVEIEIKPIPNPC--SGTRHRAMVYANVADMYRARRQDLRNAGRKLRLADASQ 815
QY 791 GIYKRGRRKDFDNLCRINKIKHSNRNLGFSNCPDCKGTGDIQTRFVVGTEWEDPEH 850
Db 816 GHYNGKANESVLD-VCDIINOYSNA-IGDSKPCNCKGSG-----FKIGTWTNIVK 867
QY 851 MRDHEVYMPRRRHCTSNLEHLOTDDHPLAGNIVYDLYVNSFLADVLLSKAYEANKI 910
Db 868 KTSYKQVFLPRRREHMCSTNLENDVGDVTNNVN-----VNNKFLVQVLLSANKQAEWI 922

QY 911 IRMYKE---KNNLGPEVTDPKHOTTICRAIRYSPADIGDILINGRDLWERNNGDMVYKLG 967
Db 923 KOKYNEBNGONNHKG-----CRALKSSFADLDIILGDTLMNDKSEQOTJÖR 970
QY 968 HLEVPFNHLSLKGKNDKYNDAPYLYKLREWMWEMANRAKWEAMKCOIKYLYKDSGH 1027
Db 971 NLYTFEIKYOKRGIDTSYTYNDGKHNOJ.REDMWEANROVWAKKCALK----- 1022
QY 1028 OSTOSSYCGYSDHTPLDLYIPOKLRMTWEMAWCYCKYOKREYKLEKCECKDKONGG 1087
Db 1023 --GERKINGC--ATPYDDYIPORLRMTWEMAWMCKOSLYDELMCKGICGMN-----G 1072
QY 1088 CTKESSGCGKCTEACMEYNDIIGLMEQNNIISDKY--KELHEQOMSVNSGIEASST 1145
Db 1073 ICKNVKDDCAKCTEACEYTKIOPMKDOMEKLELEYALSYLH----- 1115
QY 1146 AKNHIDR-----ANIEFLSELVQO--NGKSKNSTSPESAVIGNNTYENVGAY 1193
Db 1116 AKNDSRMARGTDPDYQOVVHEFKELQEAIKSSTKRPRSTDAITTDPTTYSTAAGY 1175
QY 1194 LHPTGNFDCOSQNEFCDEKSDKNEKYAFRDKPODHGACGCKSGSKPTVOIKTKKK 1253
Db 1176 IHEBIGNAGQIOKHFCDDKMD---KYVFEKPRDHBEACCTENVE-----KPK 1222
QY 1254 AEEKOTE---CKTVNDILKENOGKQVEOCHPRKKNNGYPD---WQC---GNINLEDP 1303
Db 1223 KEEDIDKLDVCAIVTALTOTDNLTKA--CQOK---YGHPRHMGKCISETTKSSDSG 1277
QY 1304 RVCMPRRORCLVHFLAN-----DNEIKKLOSOVNLKKA 1337
Db 1278 SLCPVPRRRKLYTPLYTKNAEATEBPTSPQAGEATLTPPATASQAKGSL--LTKA 1335
QY 1338 EIKSAAAEFFESYTYKSDGBSENEIDKLEKSKTIPPAFLSMFTYFGDYRDLFGT--- 1394
Db 1336 FIOSAAYEFELFMHKYK--MDNNGDAEDKLGKEIPEEFKROMFTLADYDIOICGVKMD 1394
QY 1395 ---DISKGHEGSKLKE---OIBSLFNGQOKSPNG-----KITQEWMTESHEI 1438
Db 1395 VIKALEASDKNNGNKKIKISDKIEIILKQSGSKLPGGLPTPNVKNKPTWMDONAKHI 1454
QY 1439 WEMALCALV---KIGAK--KDDFTEN-----YGYNNVFS----- 1468
Db 1455 WGMGMCALYKEDTGAJAKGTSITODPFAVYKLMNDGKPKEDYDKIYIVISSVPSN 1514
QY 1469 ---DKSTLEEFARPOPLWTFEMYDDCYTPOKYLKQVECK-----SNDOL 1515
Db 1515 VPSGDTKLEEFRRPTFRWLEMEGEFECKRKDKLEIKECYGLNRYGHIYCSGDGH 1574
QY 1516 KCD-----TECNKCEYVYKYM--KKKEMIPQDKYKDERDKKRPDRÖHI 1559
Db 1575 ICEKTDTSRNNTFIDLHCPRLKECITKYRKWIEKKEKEFNHÖKNNYEKEFN--DLKEX 1631
QY 1560 GVMWTDYGTANATDYLNRKFTASGCD--KPGASVYQNR--IOLLEKQAYYADKHC--GCTR 1616
Db 1632 G-----YSEFN-----NPLASLNHCHGHBNRDKKIEFNHMTKTFGSEYCKAPV 1679
QY 1617 FIENDKTYNISCKDKCKGLVEANTGAIKWONKGPNNYNNLKELTEVLPSPRLRCF 1676
Db 1680 YGVCKN-----KNGECETIKHTDLNG-----QNDNNYTDIKAYLVD----- 1716
QY 1677 HALGNTATDPVKNDBNGLRKLMEVAATEGYNLQYKKEKKE-----KIKTSPAUK 1729
Db 1717 ---RKGESNDELKACNNVNTSLFKDSSV-----QYMKCQKKNVDCIINFDIDIDK 1767
QY 1730 YSYEVPQASAKYSPYDLRDIILIDINLEBKÖTTEENLKIKFNKGTSVGKSGDSTTGN 1789
Db 1768 Y-----MEFVVEFQRMRLRYFVHDYNNILKDKIKPCIK-----TKDEKSNKJINGC 1811
QY 1790 PGSTARFEFWEENECEYWNAMICGYKRGDRDNGS-----NSARDEDELKKGGSV 1839
Db 1812 KGRLECYKMKLQKQDKEMKNIKDIEKKNKSLYGIPIHWMVKSIFYEVDLYFDKDKRKAÖV 1871
QY 1840 PSDDDYPMKKNRDEGTAYOFLRWFEMWGEDFCKHKEKELEKLVGACNDYTCGDNEDKRRK 1899
Db 1872 IED-----ENERKKIMGCTDGEVCTNEETKEK 1899
QY 1900 CTDACTQYKKEFISEMPOYEKOKIKYGENKDKIYSHHPAKAEDARELDQKLICEN 1959
Db 1900 --DFINILKIKLOEKIESQOTQHNPNNG--KTYPCDEIPHSD-----ETLDQÖTPTTDD 1950
QY 1960 KSGDEYK-----CMKDVSTÖRLTDGNSQMPASIDDEFEVEGKCNQVPRGPRVRE 2014
Db 1951 DMSDKLYTRKPPCEPRVDYETKET-----EKRP-----VLPGP----- 1984
QY 2015 TSPSPVLSISKATASKKEAKTAPITKQPKVYENLITPMAQÖTRRAAQÖTRKRTSTAYT 2074
Db 1985 ----- 1984
QY 2075 TESDVGTVKAILSNKPSRGGIEGCPNTY---GQYPRMGCIYKSKENENGICMPRR 2131
Db 1985 ---DACEIYGEIL--NGQÖTKKIEBCNTRYPTKNDYQWNC--TDKVINREBSCMPRR 2039
QY 2132 KRLCINNIOYLYENENKRDNDIEAFITKCAIETÖTILMY-----ITEN 2177
Db 2040 ÖKLCIHNLEHL--SEKATELELRKAFIECAAIETFWLMDKYKEDKDEKKEGGISDD 2096
QY 2178 P--AAENELÖNGTIPDEFKRIWYTYGDKDMFEFGTDSINDK--KITVYNSTYTLINENK 2235
Db 2097 PDDPKKLEGGITPDEFKROFYTGYDRDLFGTIDISKHGESALGKIDSLEFKNGDÖ 2156
QY 2236 KKÖDKKDEBELKIEWENKKNKFIWEGMITYGLTYHLTDENEKIRIDNY--QYNDMTKLPS 2294
Db 2157 KSPSGKTPLYE---WMNDYCPDIWKGWCVGSHHIKNGN--KEQJRNKNTDNNKTKTSSK 2211
QY 2295 LEEFYKRPÖPLRMTPEMAEEFCNKRQEDLKLLEGCKEYEGNSNDGTQÖCAACVYQ 2354
Db 2212 LEDPRSRQÖFLRMTIEWDÖCÖREYVAKINÖLKGECNEEYEGSÖENKKECKNACEAYK 2271
QY 2355 NFIRKWKTEYERORKEFKKDKGRKYNDYPTSTEDIEATCAHEYLMMKLKEICGNKDCS 2414
Db 2272 SWLDMKMDQYEQJAFKDKKKKF--DÖTSAEDVANAAYSVHYHLODEBLKNTLTÖKDCÖA 2330
QY 2415 CMÖKPSQLPKTIQÖSÖSSDANDPESLDYVEEF--NKCEPELSKSGSMITHKITEPK 2473
Db 2331 CMERPSAQ---DETELLGNGYPEAMDYPPKIGERCÖK-----AIPS 2371
QY 2474 IPANCVEKAAVYLSKEAENNDITLKEKEFPIESTKESKSNSTNNPCDPKPKYAPDK 2533
Db 2372 EPMSCVQÖIAHLKEKAEKNVKIYECS---IKÖTPAK--SKNDCTKID--EALIGDNGSK 2424
QY 2534 YIGR-----RNPCEMRBE---NRFVYDEMKCYKNSFYÖEKRRVCPPRRHHCLRN 2583
Db 2425 IINNSIIDSTFASNCQÖSEKDDADRLKIGKÖWF---NKINGETKLYVPRRDKCÖFND 2481
QY 2584 LDEIKIERLDSNYLTMVTRTANEGIDILKNPNSENGAMPIDDMTKYSFADIGDY 2643
Db 2482 LKNQFMEVODSNSLEKIDÖYAKNBSIDILKLNLPDÖNAFESIDÖAMKYSPADIGDI 2541
QY 2644 RGTDMLRIGGLPPEVILKLYVEEYLYGKWRNKNKGNKKNYDQTRSAWMDNRDÖIK 2703
Db 2542 RGSKID-----PTNNKRIKELÖKIFKQJÖDÖNALSÖK--ELPELRKEMWÖANREYVN 2595
QY 2704 AMTCAPEDAKLPRKGMÖFERYITLÖD-----KÖGHÖDPPVDYIFÖRFMTWESÖ 2758
Db 2596 AMTCVAPNDHLKKNKNNPNKQIIASQÖTQÖTKSHDSEPPDYIDYIPERYRFLÖWSE 2655
QY 2759 YKCKALMELEKFKKSCDHC--KTSDRCKNDYDNECÖCKTQÖEKNVLYWKSIFDQ 2817
Db 2656 YKCKALKEKNDKNDCKSCISGATCEKEDEKCKEÖNDCKEYKKNVYDQWÖSEFDO 2715
QY 2818 SNKYKELYÖPIYTKISTYDHYÖN--FVÖKLTFKSECS--VESFSEYLETSGKLYNRYKN 2874
Db 2716 NÖLYKRLYÖDÖTRHGPSTARRNPSIEFÖKL---EDSCNDPYSADKYLDISHCIDYKFS 2772
QY 2875 ENDGSSNITRYAFEEIYPRKYKACGCTILPSKNPLDNCPTDÖKKDCKÖLÖTFÖPÖSKNDY 2934

```

Db 2773 EFD-----SRESNVAFSPPKRYKENCCKV-----NTFP-----SNND- 2806
OY 2935 DNNLNNMVAIVLNNSSDDNKGVLIPRRRHLCRPTTAVNKKGDKELKKLLTSASFQ 2994
Db 2807 -----PK-----S 2809
OY 2995 GOLLQOKYSEELCEPEAKYADYSDLIKGTDMMDTSLSEKIKKIFETSEATENRKT 3054
Db 2810 PSLGPRSP----- 2817
OY 3055 WNNNRROIWHAALCGYKATSKVTLDEGCOLPKDEETNOFLRMLEWAKOACEKKHV 3114
Db 2818 -----FLPK----- 2821
OY 3115 SDSLTKCPRSEDNEPEASELIRPOCONDRIKYISLNLINKMENLIKYYKOLKDOSS 3174
Db 2822 ---KPK-----MKFPRIGIGVJHPFINM----- 2842
OY 3175 GNIDKRPSEENVQSYIKSDSCALELNDINEIVGTCKNNENREKREVLKILYPGIYFE 3234
Db 2843 -----VADPIIHETVAKTFNNVPOF----- 2864
OY 3235 DEHKHNVLDGNIKREBEQTVRPKALYFETPHVDSFYQAFVSTHRYAOYDPKNILKSI 3294
Db 2865 -----HINPKDIDVAPP-----TKNI-----LNEVLPSAI 2889
OY 3295 SVAIVSALGLALHPRMKFKSSVDLRLILNPOGEYGMPTLESKNRYIPYSGPYKGR- 3353
Db 2890 PGIALLAGSIAFLFLKKTTHPVDFSVINIPKSDYDIPTLSPRRIYPIYSGKRGNG 2949
OY 3354 YTIYVE--GDTSGDEBKMYMDLSSDIT--SSESEYELDINDIYVPGSKYITLIEVLE 3410
Db 2950 TETLKEIVTGDGXDYH-----SDIYSSSEYEELDINDIYHVLINILTLIEVLE 3003
OY 3411 PSKR-----DIPSD--DTPSN 3424
Db 3004 PSGKLSGNITPISGKNTPSD 3023

RESULT 4
T28432
variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
N.Alternate names: erythrocyte membrane binding protein 1 (Empl)
C.Species: Plasmodium falciparum
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C.Accession: T28432
R.Su, X.Z.; Healtwole, V.M.; Wertheimer, S.P.; Guinel, F.; Herfeldt, J.A.; Peterson, D.S.
182, 89-100, 1995
Title: The large diverse gene family var encodes proteins involved in cytoadherence an
Reference number: 220487; MUID:95330813; PMID:7606788
A.Accession: T28432
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-3078 <SU>
A.Cross-references: EMBL:L40608; NID:9886374; PID:9886375; PIDN:AAA73396.1
C.Genetics:
A.Gene: var-1
A.Introns: 2611/3

Query Match 18.0%; Score 3489; DB 2; Length 3078;
Best Local Similarity 26.7%; Pred. No. 7,3e-155;
Matches 1054; Conservative 472; Mismatches 934; Indels 1490; Gaps 148;

```

```

Db 109 NIKTSQGDNDGACAPRYRLHVCDDNLEQIEPIKTTNHNILVDCMAKEGOSTD 168
OY 187 HP-HKGT-----SDACTALASFPADIGIVRGIDMKPVNVD-----KYETGLREVFKKI 235
Db 169 YPKYATYGDSPSOQCTMLARSPADIGILVRGRDLYLGNPQELKQROLENNLKTIFGKI 228
OY 236 HD-----GMEDEVKNDINPDGSGNTYKILREAMVNVNKKRWEALITCDASYSGFPMOSEST 292
Db 229 YELKNGAEARLYND--PE-----FFKLREDMWTAANRKYTKATYCNV--WNTYF--HATCN 280
OY 293 PLESNPKCGHKQKGVPTNLDYVPOYLRFWDEGEERCRNRNLIKLVKVSCH-----NDKE 348
Db 281 GERITGTCACNDQVPTTYDYPQIHLRFEEBAEDCRKKNKKIDYKKNCGKDEKED 340
OY 349 RLCSHNGHDCSTTTIWKKILHLNDKCTDCSTKCYFEVWLGNOQEAFFKORKEKEYO 408
Db 341 R-YCSHNGYDCEKTKAIGLKRYGKCICSLACNPYVWMINNQKQFPQKQKKYDEIK 399
OY 409 SY-----LSNDKFFVNNINSE--YKQFYELKEIYQIYAINDPFLNLNDEGKYC----- 454
Db 400 KYENGASGGSROKRDAGGTTTNYDGEKKFYDELKSEYRTVDKFLKLSHEICTKYK 459
OY 455 --KGLPGEKDIPTNSA-----DDKGIYRSEYCYQPCQCYKCDGIXYTH--KSDND 504
Db 460 DEGGTIDFKNVNSDSTSGASGTNWSOGTFYRSKICQPCPYGKYA-----KVNNGSSNE 515
OY 505 RERVVN-----EDYKPPYGVKPTNITVLYSGNOGDIQXKLEFCN-----SS 547
Db 516 WEKNNNGKSKSGLYPRKDKEGTTITILSKGHNDIEKLKDKDEKNGOTINSGGSG 575
OY 548 IYKDKNN-----OKWECYKDENINCKLEONTEINNDPK-----II----- 586
Db 576 TGGSGGSGNSROLEYEWMC--YKGEDVYKVGHDEDEDEYEVNKNAGCILCIKQKNKE 634
OY 587 -----SEHFEELMWTYLLRDTIKAMDIKCTCI--NNTTHC--IDGCRNC 629
Db 635 EGGNTSEKPEIDIOKTENFEFTYVVAHMLKDSIHMKKLQRCIQNKKIKCGNNKCNDC 694
OY 630 LCFDRVVKQKEEEMNSIKKLFTRKKNIQ-----OSYYSNINLFEYGF-- 672
Db 695 ECFKWITQKDEMGKIYGHF--KTQNIKRGSGSDNTAELIPDHVVLVYNLOEELKND 753
OY 673 -----FKYMDKIDDEAKMKELMENIKRKNKFSLENNR-----YLENATELL 718
Db 754 SEDASEKSENSIDAEA-----ELKHLREIETSEDNNOEASVGGYTEQKNIMDKLL 807
OY 719 DILKETATIC-----KDNNTNEACESHNTATPCVYKPRGGO--PRNKIEAO 766
Db 808 NYEKDADLCLEIHEDDEEKEKGDNECIEGENFRYPCSGESGNKRYPLANKVAYO 867
OY 767 YKRSAYEBARNGLIKLKAHEGILYKRGGRKDFKDMICRIMIKSNRNLFSGPCD 826
Db 868 MHKAKTOLASRAGSALGDSISLQFKNGRNGSTLKGICIKINENYSDNSGNSGPGT 927
OY 827 GKGTGDIOTRRVVGTEWYEDREHMKHEDVIMPRRHICTSLNHLQTDHPILNGNI 886
Db 928 GK--DGDHGGVRYMRTIGEMENISGKQFSYKNVFLPRRHEMCTSNLENDV-----GSV 980
OY 887 V--DDLVNSFLDVLISAYEANKIIRMYKREKNLKGPELVDP--KKOTITCARIRS 942
Db 981 TKNDKASHLSLGVOLAAATDAEITIKRYKDNNI-----QTDIPQOKQDEAMCRAVRYS 1036
OY 943 PADIGIIGRDLMERNGDMVLOGLHLEVFNGIHSLKG--GNDKYND--APRYLKL 998
Db 1037 PADIGIIGRDMWDBDKSSTJMETRLIYFNKIKKHGINDPNRYTDESCKRPAYKYL 1096
OY 999 RENWWEANAKYWEAMKCDIKYLKDKSGHOSSTOSYSGSDHTPLDDYIPQKLRMTENA 1058
Db 1097 RADWWEANHQAWEAMK-----ATKGIIC--PGMPVDDYIPQRLRWMTENA 1141
OY 1059 EWYCKYQKEYLKKECKEKKDKNGOGCTKESGCTKCEACGEYDIIGLWMEON 1118

```

Db 1142 BWYCKASOEYDKIKKICADCMKSGDK -CT-OGDVDCGCKAACDKYKEIEKWEOMR 1119
QY 1119 IISDKYKELHBOAMSVNSGSIENASTAKNHID--RNVLEFSELYOONGK-----SNK 1171
Db 1200 KISDKYMLYLQAKTSTNG-----RVLGLDDDDPDYQOMWDLPLPHKASIAARVLKRA 1255
QY 1172 SGMSDEAVVIGTWYENVAAYLHIDGNFDCOSONEFCEK-----SDCKNEKYAPR 1225
Db 1256 AGSPTEIAAAPTTPYSTAGYIHOELGYGCOEOTOFCEKKGANSTSTYKNEKYEYTK 1315
QY 1226 DKQODHDGAGCSGSGSPFVQJTKTKKKAEBKTEKTYNDILKENDKQYVEDCPKPN 1285
Db 1316 QPPEVATACDCINRSOTE-----EPKKKEENVSACKIYEKILEGNGRTYGECPKES 1371
QY 1286 SNGYPRMOC-GNINLVEDPVCMPRRORCLCVHFLANDNEIKLOSQVNLKEAFIKSAA 1344
Db 1372 ---YPMWCKNNIDISHD-GACMPRRORCLCYIYHESGOTENIKTDMLKOFITAA 1427
QY 1345 EPEFSWYXXSK-DGEGNLDKELKSGKIPPAFLRSMFTFGDYRDLFGTDSKSGEG 1403
Db 1428 EYFLSMQYTKSKNDSEAKTIDR---GLIPSOFLRSMFTFGDYRDLCLNTDISKQNDY 1483
QY 1404 SKLEKIDSLFKNGDQSKSPNGKTRQEMWTEHSEHEWEAMICALYKGAKKD---FTENY 1460
Db 1484 AKAKDKIGFEFSMDGSKSPSGLSRQEMWTKNGEYIKMGICALITYVDITDMKRIKNDY 1543
QY 1461 GYNNVYFSK-STYLEEFKAKPOFLMWTETWYDYCYTRYKYLKVOEKCS-NDOLKCD 1518
Db 1544 STDKNQSONGSPSLEEFKAKPOFLMWTETWYDYCYTRYKYLKVOEKCS-NDOLKCD 1603
QY 1519 T---BCNKCEDYVYKMK-KRKEMIPQDYRDERDKKRFDOHIGVWYDVTGTINADY 1574
Db 1604 DAKHRCNQACRAVOEYVEKKEFGSGQTNNFLKANKVOPDEYEGYEYKD----- 1654
QY 1575 IAKKFTASGCGPSSAVQONIOLEKQAYYADAKHCGCTFIENDDKYTNISSKDKC 1634
Db 1655 ---GYOP---IOGNEYLLQK---CDNNK-CSC----- 1676
QY 1635 GLVKEANTGAIKMNKGNPNYNNLKEJEDVLFPSRRLRICHALDNGTYTDEYVDEGL 1694
Db 1677 -----MDGNVLSVSPKEK--- 1689
QY 1695 KRRLMEVATEGYNLGOYKKEKKEKIKTSDAHKYSYEPFPCSAMKYSFYDLRDIIGI 1754
Db 1690 ---PRGXT-----AHKYP----- 1699
QY 1755 DNLDEKQTEBNLKKITNNKGTSGKSGSDSTGNGPSTARKFWENNEKCVWNAITCGY 1814
Db 1700 ----- 1699
QY 1815 KRGRDNGSGNSARDEDLKKGSGVSPDDPYPMGKNRDEGTAYQFLRWFAEMGEDFCXHK 1874
Db 1700 ----- 1699
QY 1875 EKELEKLVGACNDYTCGDNEKRRKCTDACYQYKFKFISEMFPQYKQIKKYGKNDKIYS 1934
Db 1700 -----EKC----- 1702
QY 1935 EHPVAKAEADAREYLDQKIKCENKSGDCYKCMKDYSTORLTDGNSQNNPASLDEPK 1994
Db 1703 -----DC----- 1704
QY 1995 EVEGCMCOYRGPBRVRETSPRVSLSIKATASKKEAKTAPPTKQPKYENTITEMRA 2054
Db 1705 -YGCK--HVPSTP-----PPPVOPQ----- 1723
QY 2055 QTRRRRAAOQTRKRTSTATTSTESDVGMVAKAISNNKPDSSGKIEGCMPTKYQY--PKWG 2112
Db 1724 -----PEAPRYTVYDVCSIVKTLFK--DTNNFSDAGLK-YGKTAPESSMK 1764
QY 2113 CI-----YKSGKENENGICMPRRKIKLGINNIOYLANET-----EKKD 2151
Db 1765 CIPSDYTSAGAGATTGKSGSDSGICIPRRRRRLYGVKIQ--EWATALPQEGGAAPSHSRA 1822

QY 2152 NDIKFAICAIETOFMLKTIIB-NPAENE-----LQNGT 2188
Db 1823 DILRNAFIOSAIEFFELMDRYKEKKPOGDSQOALSQLTSTYSDEDEPPDKLLQNGK 1882
QY 2189 IPDEFKRIYVYGYKDMFECTDISNDKKITVNSYTLINENKOKKODKDEELK 2248
Db 1883 IPPDLRLMFTYLGIDYRDLVHGNTSDSGNTGNSNNNIVLEASGNKEDOKIOEKIEO 1942
QY 2249 I-----FWENKKFIWEGMITYGLY-----HLADNEKEKIR 2280
Db 1943 ILPKNGTPIVYKSSAQTPDKWNEHAEISIMGMICALITYEKNEDTSARDEKTEKD 2002
QY 2281 DNY-----QYN-DMTKL-----TPSLEFYKRP 2302
Db 2003 EYIEKFFSGTADKHGTASTPTGYTYKYDYKVLKEDTSAGKTPSASSDTLLSDFLVR 2062
QY 2303 QFLRMFTEMAEEFCNKRREOLKLEAGCKEYE-----CN---GSN 2339
Db 2063 PYFRLIEMGONFCCKRKHKLQAIKHEKVEENGGSGHRGITROYSGDGACEMLPKN 2122
QY 2340 DKTQOE---CAEACVYQNFIKKWTYERQBEKFKKD---GKKYDYSTERDI 2390
Db 2123 DGYVDLEKPCAPKCSYSRWIESKGEFEKQKAYBQKDCVNGSNKH 2174
QY 2391 EKATCAHEYLMKLELGNKDCGCMQKPSQLKTTQOOSSPANDMPESLDYVPEEPN 2450
Db 2175 -----KFCETL-----TSSAKDFLTLG----- 2195
QY 2451 KCEPELSKSGSMI-----HTKKITEP--KIPMNCVEKAAVYLSKEANNMDITLKE 2500
Db 2196 -PKPNVNEGKITIPDDKTFEKKTKD--CDPCLKFSVC----- 2230
QY 2501 KFIPIESTKEKESNMTNNPCOPKPKPYADKITYGRNCPENRENRFYVYEMKCYN 2560
Db 2231 ---KQECBNS-----KGTDCRN-----KN 2247
QY 2561 SKFYOEKRRVCVPPRRHMLRLNDEIKIRLKDNYLLKMWRTARNEGIDILKNFNS- 2619
Db 2248 S-----IDATDIENGVDSTYLEMRVSDSKS-----GENGD 2278
QY 2620 --ENGCAINPLCTMTKSFADLGDYRGDTMLRIGYLPVEIKLYKVEFYIGKWRKN 2677
Db 2279 GLENAC-----RGAG-----IFGCI----- 2293
QY 2678 KGRKYNDVQTFPSAMWANDANKDIKRA-----MTCKAPEDAKLFRKGRDGERITLLO 2731
Db 2294 -----RKDEMKCRNVCYGVCK-PENVNGEAKG----- 2321
QY 2732 DKCGHKDDPPVDYIIPQRFWMATEWSEYCYKALMELEKPKSCDHCQKTSDRCKNDYDEN 2791
Db 2322 -----HIIQIRALVAKWEIF--FEDYKTKH-----KISHRIK--G 2355
QY 2792 KCBOCKTRCOBYENFYAKKSLPDIOSNNKYLEOPITYKISTYDVHONFYOKLTFKS 2851
Db 2356 EISPC-----IKNCVEKW--VDOKRKKEWKEITER-----FRD 2385
QY 2852 ECVSESEYELIHTSKLYKFNENDGSSNIRTYAREETPK-----SYKRA 2897
Db 2386 Q-----YK-NDNSDDNVASFLETLIPQITDANAKKVIKLSFGNS 2426
QY 2898 GSGTILSKNLDNCPDQNDKCGCKELQTFPCSKNDYDNNLJMNWAVIYLVNSSDDNKGLV 2957
Db 2427 CGGSASA-----NEONKNG----- 2440
QY 2958 IPRRRHLCRPIITAVANYKKQKKEILKLLKLLSABSOGLLOOKYKSEBELCFEAMKYSY 3017
Db 2441 ----- 2440
QY 3018 ADYSIIKGTDMWDSLSERIKKIETSNENATENKRTWENNRRQIWMHMLCGYKIATSK 3077
Db 2441 -EYKDAI--DGLKLLKDKI----- 2457

QY 3078 VTLDGECQOLPKDEETNOFLRLLEMAKQACKKKHSDLSKTKC-----PRSNEDNFEA 3132
Db 2458 -----GEC-----BKKHQTS-DTECSPTPOPTLE----- 2483
QY 3133 SELLRGCGCONDIRKTYISLNLKNTMENTNRYKOLKQDSSGINDKPSSEENVQSYTKS 3192
Db 2484 -EHL-----DDLETBAKKNNMPKICEV-NLTAQODE-GGCVAPENSEEPA-----A 2530
QY 3193 KDSOCALENDINEIYGTNNNNNEKEVYKLLPGLYVEDETHKNHVLGDNKEEQ 3252
Db 2531 TDS-----GKEPDEQTPYLK-----PEEA 2550
QY 3253 TVRKALYFTTPHYDSTQAPLESTHRAVQYDPKN-----DLKSSIVYVYALG 3303
Db 2551 VPBP-----PPPOEKAPAP-----IPQPPPTPTOLLNDNPHVLTALVSTLAVSVG 2599
QY 3304 L-----LALHFMKKKSSV-DLRLINIPQEGEMPLESKNRIYPRSPYKTYIM 3358
Db 2600 IGRATFYFLKKTSSVGNLQIOLIPKSDIDIPKLSPNRIYPTSGKRGKRIYTL 2659
QY 3359 EGDY-----SGDEKYMMDLSSDITSESEYEELDINDIYVPGSKYTLLEVLPSKRD 3415
Db 2660 EGDGTDGTYDHY-----SDITSESEYEEMDINDIYVPGSKYTLLEVLPSGNN 2713
QY 3416 IPBD-----DTPS-----NDPTNRRTIDEMNLKDFVSQYL-- 3449
Db 2714 TTASGNNTTASGNNTTASGNTPSDQNDQNDQIPSKITDMNOLKDEFISQYLQSE 2773
QY 3450 PTEPNNTKYSADIPNTEPNNTLYSDNPEKPFITSHDLYTGKELSY----- 3499
Db 2774 PNEP-----NMGYNVDNTHHTTSH-HAVEKPFITSHDNLBSGEYNTDMFSGNNPI 2830
QY 3500 NINNSTN-----TNNDIPNARNDYRGIDLINDSLVNLILY--MMK 3541
Db 2831 NISDSTMSDLSNNSNPNKNDLYSGIDLINDALSGNHIDYDEMLK 2880

RESULT 5

T18378

variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence, revision 15-Oct-1999 #text, change 09-Jun-2000
C:Accession: T18378
R:Baruch, D.I.; Pasloske, B.L.; Singh, H.B.; Bi, X.; Ma, X.C.; Feldman, M.; Taraschl, T.
Cell 82, 77-87, 1995

A:Title: Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and
A:Reference number: Z18925; MID:95330812; PMID:7541722
Accession: T18378
Status: preliminary; translated from GB/EMBL/DBJ
Molecule type: DNA
A:Residues: 1-2924 <BAR>
A:Cross-references: EMBL:U7338; NID:9914918; PID:9914919; PIDN:AAB60251.1
C:Genetics:
A:Gene: EMP1
A:Introns: 24/76/3

Query Match 16.1%; Score 3130; DB 2; Length 2924;
Best Local Similarity 26.7%; Pred. No. 4.3e-138;
Matches 1005; Conservative 445; Mismatches 968; Indels 1350; Gaps 151;

QY 33 KSARNVLERAKNI-----RHSKYAKEHVDSLKGLTAAEFPGSPVYNNHNYYPY 86
Db 13 KDAKHALDRIGEEYKKEKVENDAEKYK-----ALKGNLOEKAGIGELASS-----PN 60
QY 87 PCNLNDEKHTN-LRYDVNLRHPCHGREGONRFEDEDESEGC-NKIRYKRNDAI-ACAP 143
Db 61 PCKLVEYTYNNRKL-----KRYPCANRQYVAFSDEYGGQCTFNRIKQSENNDIGACAP 116
QY 144 PRRRHMDKNLEALNDINTONIHDLGNVLTAKYEGESTYNNH-PHKG-----SDACT 197
Db 117 YRRLHLDYNNLEKMGKSTTK-HDLLLDVCAAAVEEDSIRKHTYTKHLEINPDIKSOLCT 175

QY 198 ALARSPADIDIVRGIDM-----KPNHDKVETGLREYFKKIHGDMEDEKNDVNDG 251
Db 176 ILARSPADIDIVRGIDM-----KPNHDKVETGLREYFKKIHGDMEDEKNDVNDG 251
QY 252 SGNYTKLEAMVNNRKNVWEALITCDASYSKGFEMOSESTPLESN-KGHNQK----- 306
Db 236 EENTYOLREDMWTANRSTWAKALITCHAGESDKFRKCCSGEWTDDCKRCKDEGKNETN 295
QY 307 -VPTNLIDYVQYLRFEMDEEFCRKNRIKLKYKQDSCRNDK-ERLYCSHNGHDTT 363
Db 296 EVPTLFEYVQYLRFEMDEEFCRKNRIKLKYKQDSCRNDK-ERLYCSHNGHDTT 363
QY 364 WKGIILHNDKCTDCSTKCVFEVWILGNQOAPKQKQEKYKETO--SYLSNDKFN 419
Db 355 RGAEIFVKGDDCHCSVACDRFVWILGNQOAPKQKQEKYKETO--SYLSNDKFN 419
QY 420 NINSEYKQFYEKLKEVQVNTDFNLNLBEGYCK--GGLPBKDTTFNSADKGF 477
Db 415 NL--YVGHFYKILKK-YFPVDSLOKLEDAICRPPVNGEKASTVDFNNEVNTFS 470
QY 478 RSEYQVCPDGYKCD-----GIKYTHKS-DNDRERVNEDYKPPWGVKPTNITVYSGNE 532
Db 471 HTTYCEACPGCAGQKENGNGGWAKKESCAKKERIKNE-----NSTDIKILPEKG 523
QY 533 QGDTQKLENFCSSTYKDKRNNQKMECYKKNENINCKLEBONTLEIND----- 581
Db 524 RSKLEKTKFCKGQKQKIK--NDIWKCHYDNDG-----TDQDITDSDNCVLGDMNLTK 575
QY 582 NPKIISPHNEFLWVYTLADPTIKMNDKLTCTIN-NTTHCIDECCNCCCFDPMVQKE 640
Db 576 EDKIMSTNAFFMVDMLDSTIKWDEHRCIKNDKGTICIGCNKCKICFOKWWQK 635
QY 641 EENNSIKLEFTKKNIQOOSYNNINLDEGYFFKVDKLDKDEAKWELMENIKRKNF 700
Db 636 TEMGKIDHRRKQDIDKDM----- 655
QY 701 SNLENNRDIYENAI--ELLDLKLEVTICKDNNTNACETSHNATTPCVRPGSTOPT 758
Db 656 ---THDPLQTLMLKDLLEIIDDY--GDANE----- 683
QY 759 KNKEIQYKRSAYEARNRGLKLGKAGHEGYKRGGRKQKDNLCIMIKHSNRL 818
Db 684 -IKRT-----EALBDA--GGGIDPALAGLYTGVAE-KDPTIDKLQHEOK-- 729
QY 819 GFSNGPCDGKGTGDIQTRVVGTEWEVDEHMKDHEVYIMPRRHICTSNLEHQT 878
Db 730 -----EADKCKTKTDTDTCPQ-----E 747
QY 879 DHPNLGNITVDLVNNSFLGVLISAKYEANKIIRMYEKNKNGKPEVTPDKHQTTCRA 938
Db 748 DRSV-----ARSESATV-----PSPADPKATEEV--- 772
QY 939 IRTSFADIGIIRGRDLMEBNGDVYKQGHLETFYFGIHSKSGKNDKYNDAFKYLK 998
Db 773 ----- 772
QY 999 RENWMEANRAKWEAMKCIKYLKDKSGHSTOSSYGSYSDHPDIDYIPQKLRMTENA 1058
Db 773 -----DANAS-----SDDE--DDEEBE-- 787
QY 1059 EWTCKVOKKEDYKLECKECKDKDNGQCGTK--ESGTCGTCTEACNEYNIDILGLMKE 1115
Db 788 -----EEEBEDEEBEAEVEQEKTEBSATEAVAPSPPTG----- 824
QY 1116 QNNTISQYKLEHQAOAMSVNSGIEASSTAKNIIDNVLEFSELYQONGSKSNKSGTS 1175
Db 825 -----DGVPASQEDDYKQCS-----VDALKGLKLDADCTLYKGT--APTS 865
QY 1176 DESAVIGTNTTYENGVAYLHDGNFDDCQSQNEFCDEKSDGDKNEKAYAFDKPDHDGAC 1235
Db 866 WKCPSPSNNTTST-----TKP-----GAA 886
QY 1236 GCKSGSKPTKVOIKTKAKAEKTECTVNDILKENDGKQVDECHPKKNSNGYPPMOCG 1295

Db 887 GTPSG-----KDT-----G 895
Qy 1296 NINIVEDPVCMPPRRKOCVHPL-----ANDN----- 1323
Db 896 SI-----CVPRRKRLVYGLHDMAGETTBAKSQETSQOKTPSGNESSPSEKLPQ 947
Qy 1324 ----EIKLOSQVNLKAEATKSAAEETFSWYYSK-----DGEg 1360
Db 948 GHPETETKEPESLHAFAVSPRLRFLPMHFKPEQWKAHGAGANGCOOTITIGTLDGG 1007
Qy 1361 NEL-DKELTEGKIIPAPLBSMETTFGDRFLG-IDI---SKHGEGSKLKEQIDS-- 1412
Db 1008 EEPDLKLTGHIIPDLROMFTYLGDIIVHTSGNEDMOIMEAIOKIE 1067
Qy 1413 --LEKNGDOKSPNGKT-----ROEMWTEHSEIWEAMLCALV----- 1447
Db 1068 OILPTSSSPSPRVOTQOSVENPRKTWNENGGKIMEGVICALYNTDPSGTAPTOI 1127
1448 -KITGAR-KDFTN-----GYNNVFRSD-----KSTLEEFARPOFLR 1485
1128 QEVRTLRBENSNNPKIPQYKYDQVKLDITSDAKTTGSPVSGEKLITPLDIFSRPFR 1187
Qy 1486 WLEWYDDCYTROYKLYKDOEKC-----SNDOLKC---DRECKK-----CEDY 1528
Db 1188 YLEMEGETCKEKKKLEKIEKECRGDRGHEHCSDGYDCTRTDADRNDKFDVLCRDC 1247
Qy 1529 VKYMKRKKEM--IPODKYYKDEBRKKRFDROHIGVAVTDTGTNATDYLNRKFTASGDK 1586
Db 1248 HICRCRYKRWIDIKFDEYHKOEK-----KYOG---EY-----DK 1278
Qy 1587 PGSASVVOANIOLEKQAYADADHGGCTKFTIENDDKYNNISKDKCKGLVEAMTAIRK 1646
Db 1279 -----LNDKSSGGDNMC-----CKDIEKHKSAVP- 1304
Qy 1647 MONKGNPNYNNKLETEDEVLFPSRLRIGCFHALDGNVTDPEYKDENGKLRLMEVAATG 1706
Db 1305 -----LKEK-----HCKNGOTSNNKNGEDOL- 1327
Qy 1707 YNLGQYKKEKEKEKI-KTSDAHKSYEVPKCSAMKYSFYDLRDIILGIDNLEDEKOTE 1765
Db 1328 -----NKLDPRKIPQFSPSTYCKACP-----VYGV----- 1353
Qy 1766 ENLKIKFNKSTGVKSGSDS--TTGNPSTARKFPMNENKECVNANMC-----GYKR 1816
Db 1354 -----NONGNKRGRGNGCTTNN-----PENKENDKGAASITSLINDGSTN 1397
Qy 1817 GDDGNSGNSASBDLKKCGSPSDDYPM--GKNRDEGTAYOFLRWFAMGEGDFCKHK 1874
1398 GATNGTGTG---TDETLKEC-----SDKYAFKGLRQEWTCOK-----KGVNQ- 1440
Qy 1875 EKELEKIVGACNDYCGNEDRKRCCTDACYOKKFISEMKPOYEKOIKKKGEMKDKIYS 1934
Db 1441 ----NLNRRVND-TYFKD-----IVNFEFQRMWRYF---VHADNIIKHKI-- 1479
Qy 1935 EHPVAKDAEDAREYLDKOLKICENKSGDCEYKCKJMDYSTORLTDGNSGNPASLDEPK 1994
Db 1480 -----DPOIKKEKQKT---EHKCI-----NGCNI----- 1501
Qy 1995 EYEGCNCQVRRGPPRVRRTPSPRVSLSIKATASKKEAKTAPTKOPKVENLTTEKRA 2054
Db 1502 ----KCEC-----VR----- 1507
Qy 2055 QUTRRAAQOTRRRTSTATTTESDVGIVVKAISLKNKPSRGISGCNPKTYGYPRKMGCI 2114
Db 1508 -----KMLEI 1512
Qy 2115 VGKKEENEGICMPRRKRLCINNIOYLYNTEENKRDNDIEAFIKCAIETQFLMKYI 2174
Db 1513 KG---NEMG-----NIK---KHYNINSDDKEI----- 1534
Qy 2175 IENPAENELONGTIPDEFKRIIMTYTGDKDM-FEGTDISNDKKIITVNSVTTLINEN 2233
Db 1535 -----IAYNKSTYVDQGLFDDYDKRAQVY----- 1560
Qy 2234 NKKROKKDEELRLTEWENKKNKFIWEGMIYGLYHLEDE-NEKKIRIDNYOYNDMTKLT 2292
Db 1561 -----BDEKERKKIW-----GCTGH--DECEKKEKENK---NITNLI 1594
Qy 2293 PSLEFPYKPOFLRWTEWAEFCNKKRQOLKLEAGCKEYEGNSNDGTOEACAVT 2352
Db 1595 SELQDKTYSQO-----NKH-----NPNKGT-----AC-- 1616
Qy 2353 YONFLKMWTEYEROREKFKDKDKKDYKDYPTSERDIEKATCAHEYIMMLKELCGKD 2412
Db 1617 ----- 1616
Qy 2413 CSCMOKPSSQLPKTQOSOSDANDPBLDVPBEFNKCEPDELSSKSGMIHKKITEP 2472
Db 1617 -----DPPSPPEEFDPLDDTDPDLDDQHTBOQPKCP-----P 1653
Qy 2473 KIPMNCVEKAAVYLSKEAENNDITLKEK-----FPISTK-EKESK 2514
Db 1654 PPPMTCVEKIAKELVEABGAKINNELKNGDENGKCNVKKNGAVIGESCKFEQTYE 1713
Qy 2515 NSWTNNNCPDPRKPYAPDKYIGRRNCPENREBNREKVDYEMKCYKNSKFTYOE-KRVCVP 2573
Db 1714 NSVNIN-----NKKDQNGNERKIGQKM-----NRYIGTIRKDLICIP 1752
Qy 2574 PRREHMCRLNDEIKIEKLSNITLKVVRRTARNEGIDITIKFNSGNCAMPIDCTMK 2633
Db 1753 PRREHMCRLDLSMIGRTTISOSSALKKIOEPAKSERDITIRKILEGNSCEHRIQAMK 1812
Qy 2634 YSFADLGDIVGTDMLRIGGYLPPVEIKLYVFEYIYGKMN--KN-----KGNKNNDVQTF 2689
Db 1813 YSFADLGDIIINGRDMKNNKSKQGLQKLEAFINITYNKLNDNNKYEKDKPKYLOI-- 1869
Qy 2690 RSAMWDANRKOIMWAKTCADEAKLFRKGRMD-----GEBRITLIDCKGSHKDDPVD 2744
Db 1870 RSDWWDANRKHIMWAMCNAPDADAKFLKKNPDNSGSSSSKIGMTHTSGNGGYOKEPPDYD 1929
Qy 2745 YIPORFMWTEMSRYCKALMEELKFKKSGDHC-KTSDCKNDYDNKCEOCKTRCOEY 2803
Db 1930 YIPORFMWTEMSRYCKALMEELKFKKSGDHC-KTSDCKNDYDNKCEOCKTRCOEY 2803
Qy 2804 KNYVLKMSLEFDIOSNRYKLEYOPY---TKISTYDHVONFOYOKLTFRSEK-SWESF 2858
Db 1990 KKLHNMKLGFD---KYKEIYNE-IYNNKDSKINSNEYKFKLEK---KDKCKELNNS 2041
Qy 2859 SEYIHTSKCLNTYKFNNDSSNRTYAFETPRSYEACSCILPSKNPLDNCPTDO--N 2916
Db 2042 DKCIDEATHCTKIKFS--NSEKNNNNYAFKNPPEYERKACDAP--DPLDNCPRKSATY 2098
Qy 2917 KDGCCELQTFEFGSKNDYDNNDLMMWNAVLYLNSDDKRGVLIIPRRRHLCRTRDITAVNR 2976
Db 2099 EKACNTLLPTRLCKSKTFNNDDSDWDSFYQTSRDTGVLVPRRRROLKNTTY-KLR 2157
Qy 2977 KGGR-ELKKKLLTSAFSQOLLGQYKKSSEELCFEAMKSYADYSDILKGTMDMTJSL 3035
Db 2158 SIEKIDDFKALMTSAVNEKGLCELYKKDRDVTLOAMKYSFYDGYVAGTDLISTAPL 2217
Qy 3036 EKIKKIF-----ETSSEATENRKTWENNRRROIWAMLCGYIATISKYTYLDSGQOLK 3089
Db 2218 DKLTAKLNTLKGDNTEIREDKGMWTEKNTRTVWAMLCGYAAAGKI--ERRDCSLP- 2274
Qy 3090 DEETNOFLRWLIMWAQACEKKAHVSLSLTKCPRSNEDNFEASELLRPGCO---NDIR 3146
Db 2275 DDNTHQFLRFRMSSEHFCARQKLFNEYKREC--ASACIIIEYGITDPPVCEACTQYR 2332
Qy 3147 KYISLNLKNTYENNLKIKQKLDOSGNDINKPSE-ENVOYSIKSKDSOCALDELIND 3205
Db 2333 DYTTRKIO-----EYRLNLYOYNTNFENKKAELVAKAEYFMDK---CNDKNCJLS 2379
Qy 3206 EITGTGRNNEENFEKFLVLLKLYGLYFVEDETHKNHVLGNGIKKEEQGYRPAALAFETPH 3265
Db 2380 KYI-----DIEKKWAMYSF---DDNLLKNCICRQIKPKRP---PKVAPREEH 2424

QY 3266 VDSFYAPLEFSTHRAVQYDP-----KNDLKSSISVIVSALGLIALHMKRKRKSV- 3318
 DB 2425 TSEEDQTPPLPPLPKPDLPPEAPEEPNRDLIEKTIPEFGIALALGSLAEFLKAKTKSSVG 2484
 QY 3319 DLIRINIOGEGMPTLSSKNRYIPYRSRPGYKTYIWEQDT---SGDEKTYMDLSS 3375
 DB 2485 NFEQILHIPSDDLPDKLSPNRYIPYTSKGRKRYIYLGSDSGTGDYH----- 2538
 QY 3376 SDIT--SSESEYEELDINDIYVPGSPKYLIEVLEPESKRDIPSD--DPFS-----ND 3425
 DB 2539 SDITSSSESEYEELDINDIYVPGSPKYLIEVLEPESGNTTASAKNTPSDQNDIQND 2598
 QY 3426 TPRTANFIDDEWELKHDFVSOYL---PTEPNNTKASADIPNTEPNTLYSDNPEKPF 3482
 DB 2599 GIPSSKITNENMTLKDDEFISNMLQNEPNER--NMIGYVVDNNTHTPT--SRHNVKEKPF 2655
 QY 3483 IISIHRODLYTGKEISYINIMSTNTNDIPMARNDSYRGIDILINSL 3530
 DB 2656 IMSHRDLXSGEESYINVM--VNDLPIISAKNGYSGLINDSL 2700

RESULT 6 T28626

variant-specific surface protein 2 - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 J:Accession: T28626
 R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Gulnet, F.; Herfelde, J.A.; Peterson, D.S.
 Cell 82, 89-100, 1995
 A:Title: The large diverse gene family var encodes proteins involved in cytoadherence an
 A:Reference number: Z20487; MID:95330813; PMID:7606788
 A:Accession: T28626
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2664 <SD>
 A:Cross-references: EMBL:L40609; NID:g886376; PID:g886378; PIDN:AAA75398.1
 C:Genetics: 2197/3
 A:Note: var-2

Query Match 14.0% Score 2714.5 DB 2; Length 2664;
 Best Local Similarity 23.9% Pred. No. 1e-118;
 Matches 903; Conservative 355; Mismatches 879; Indels 1605; Gaps 123;

QY 33 KSARNYLERAKNIRHPSYAKHEVDS-----LKGDLTKAEFRGSPSTPVNKHNYYP 85
 DB 17 RSAKHLLDSIGKIV-----YDKVHGADALQPSNGKLGITSLAEKAPBEKQTSED----- 67
 QY 86 YPCNDLHKERTNR--YDVYNLHPCHGREGONFDEDESECG-NKIRNKKRNDALACA 142
 DB 68 -PCDLNHEHTTYTSGYDKEN--PCKDREVERFSYEGAECDKSKIRGSNSKDG-ACA 122
 QY 143 PPRRRHMCDEKLEAL-NDITONIHDLGNVLTAKYEGSI-----VNNHPHKGS 193
 DB 123 PPRRLHLCDOHEHIKDKTR--HNILADYCEAAKFEASLEKRYQOYLN-----S 174
 QY 194 DA---CTALARSFADIGDIVRGIDMEKPN--VHKVETGLREYFKIHGMD-ERYND 246
 DB 175 DVAINICTELARSFADIGDIVRGIDMEKPN--VHKVETGLREYFKIHGMD-ERYND 246
 QY 247 YNPD--GSGYTYKLRAMNVRNRYKWEATITCAYSKSGTYMOSSESTPLFSNPKCGHKQ 304
 DB 235 YKDDDDKTYKTYKLRAMNVRNRYKWEATITCAYSKSGTYMOSSESTPLFSNPKCGHKQ 304
 QY 305 GKVPYNDYVQYTLRMEDEGESEKRRNKLKRVKDSGNDKRELYCSHNGHDCITTM 364
 DB 293 NDVPTFDYVQYTLRMEDEGESEKRRNKLKRVKDSGNDKRELYCSHNGHDCITTM 364
 QY 365 KKGILHLDNKCCTDSCRKCYEYVWLNQOAEFRKQEKYKEIYSLSNKRVNNINSE 424
 DB 353 VIGHNVIGSECSKCYEYVWLNQOAEFRKQEKYKEIYSLSNKRVNNINSE 424

QY 425 YNKFTEKLETOYATNDPFLNLNBSGKYK-----GGLPEKDDITFTNS 469
 DB 413 YNKFTEKLETOYATNDPFLNLNBSGKYK-----GGLPEKDDITFTNS 469
 QY 470 ADDKGIYFSEYCOVPCDGVKCDGJKYTHKSNDRERVANNELKRPNGKYPINITYLV 529
 DB 473 NNSNKTFSHOCYCEGCGGCVLIGLWEM--KEKNKGCGKGRNTPKGTNHNIVPLSF 530
 QY 530 GNEQGDITOKLEFNCNCSSTYKKNNOKECYKDEININCKLEQNTLEINDPKI-ISP 588
 DB 531 GDBHKEIIEIEQFCABNSDSSSKLTFQWKCYYGDKREYVCTLENNKNSSEDEBELQKPF 590
 QY 589 HNFELWATYLLDITKANDKLTCTINNTT--HCIDECNBLCLFDRWYKQKEENNSIK 647
 DB 591 HNFELWATYLLDITKANDKLTCTINNTT--HCIDECNBLCLFDRWYKQKEENNSIK 647
 QY 648 KLFTEKKNIOQSYSNINLFEYGFYVMDKLKDEAKMELEMNIRKKNESNLENNR 707
 DB 651 KHKF-----TQDFSTFGN--NTDYLELVNID-----ELFQDITEATGNSQKIOGIR 697
 QY 708 DYL-----ENALIEDLDLAKETATICKDNNTNEACETSH-NATTNPGVKKR 752
 DB 698 DTLAKKKTOAADATBQKNTIDLLFEYDSEAEKCK--KIQECCQPKPKTYVANNPCY--- 752
 QY 753 GGTQPTKNTKEIAYQYRKRAVEEA--RNRGLHKLKGAHSGIYRGGRKRFKNLCRIMI 811
 DB 753 GNTYTDALACKVAKLOEQEKEQLDRNDSRAKLANASQKYSNOGDDPDKKNLGITO 812
 QY 812 KHSNRNLGFSNGCDGCGTGDGIGTQTRFYVGTWEMVDEPHKRDHEVIMPPRRRHCTSN 871
 DB 813 KHSNA-IGDSKNPCNNKG-----KERPNVEKKNKGE-VKMSHTDYLLEPRKHQFTSN 865
 QY 872 LEHLOTDDHPLNGNIYVDLVNNSFLGVLISAVYEANKIIRMYKEKNKLKPEVTPDKH 931
 DB 866 LEHLOTDDHPLNGNIYVDLVNNSFLGVLISAVYEANKIIRMYKEKNKLKPEVTPDKH 931
 QY 932 QTTICRAIRSFADIGDIVRGIDMEKPN--VHKVETGLREYFKIHGMD-ERYND 246
 DB 910 RSSICRTMKYSFADIGDIVRGIDMEKPN--VHKVETGLREYFKIHGMD-ERYND 246
 QY 991 ----DAPKYTLRENNMEANRAKRYEAMKDIYLYKDKSGHSTOSSGYSHTPLDXY 1046
 DB 970 DKTNDNSKHKILKSDMWEARDOYKAMTCTPK-----NGNTQCG--ATPDDY 1016
 QY 1047 IPQKLRMTEMAWYCKVQKREYDKLKECKECKDNDGCGTRESGTGCTKEACNEY 1106
 DB 1017 IPQKLRMTEMAWYCKVQKREYDKLKECKECKDNDGCGTRESGTGCTKEACNEY 1106
 QY 1107 NDITGLMEQWNIISDKYKFLHDOAQMSVNSGSIKASSTAKNIDRNVIEFLSELYQNG 1166
 DB 1070 KTRIQPADWQNEISNRYQILYQAKIAALINGSTERTKDKDKVYIDFLQKLEBANY 1129
 QY 1167 GKSNSGTSDES-----AVIGTNTYENWAGAYLIDTGNFDDCGGSECEKSDGKD 1218
 DB 1130 G---TRPPPEAPDRPRRAATSKSDYETIAGTITQEKARTRECLQANFCNNNG---- 1182
 QY 1219 NEKYAFRDKPDQHDGACGCKSGSKPRTVOJTKTKAEKDEKTEKYNDILKENDGKQVE 1278
 DB 1183 NNEVAFSLDHEYKHAACK-----NEKKASS-----PEELGRSDSF 1218
 QY 1279 DCHPKRNSNGYPMQCGNINLVEDPRYCMPPRROKLCVHFLANDNEIKKLOSOVNLKEAF 1338
 DB 1219 DDH-----QTPREDEEVH----- 1232
 QY 1339 IKSAAAEFFSWYTKSGDEGNELDEKELKGRIPAFLSMFTYFDYDPLFGTDISK 1398
 DB 1233 -----SEEBEEDSEDEKE----- 1248
 QY 1399 GHGEGSLKQDIDSLFKNGDQKSPNGKTRQEMWTESHSIWEAMCALVIRIGAKKDDPTE 1458
 DB 1249 -----BEVEEVDGADEKA----- 1262
 QY 1459 NYGNNYKFSDKSTILEEFAKRPQFLMLNTLEMTDIDCYTQKYLKVOECKSKSNDOLKCD 1518

Db 1263 ----- 1262
QY 1519 TECNKKCEDVYKMKKKKEMIPQDKYYKEDBKRPDRHIGVAVTDTGTGNATDIYLRK 1578
Db 1263 ----- 1262
QY 1579 FTASGDKPGSASVYORNIOLLEKQAYYDADKHGCTKFTENDKYNISSKDKCKGLVK 1638
Db 1263 ----- 1262
QY 1639 EANTGALKMOKGPNNTNKLKEDLYLFPSSRLRITCFHALDGNITDPEVKENGELKRL 1698
Db 1263 ----- 1278
QY 1699 MEVAATEGYNLGOYKKEKKEKIKTSDAHKYSEYVPCGAMKYSFYDLRDLITDNL 1758
Db 1279 ----- 1296
QY 1759 DEKOTENKIKTFNNKGTSVGKSDSTGTGNPGSTARKFWMENKECWAMIGYKRG 1818
1297 ----- 1296
QY 1819 DDCNSGNSARSDBDLKCGSVPSDDYPMGKNRDEGTAYQFLWPAWGEDPCKHKEKEL 1878
Db 1297 ----- 1308
QY 1879 EKLVGACNDYTCGDNEDKRRKCTDCTOYKKTISEMKPOYKQIKKYGENKDKIYSEHPV 1938
Db 1309 ----- 1310
QY 1939 AKDAEDAREYLDKOLKIKGENKSGDEYKCMKOVSTQRLTDGNSQWMPASLDEPEREVG 1998
Db 1311 ----- 1328
QY 1999 KCCQVOPRPPRYRRETPSPRVSLSKATASKEAKTAPPTPOPKVENITTEMAQRT 2058
Db 1329 ----- 1364
QY 2059 RRAOQTRKRTSATTTESDVGTMWKAIIISNRPDSGEGCNPKTYGOYPRMGCIYKGS 2118
Db 1365 ----- 1395
QY 2119 KENENGICMPRRKICLINIOLYVETENKRNNDIKFAIFICALIETQFLKTYIE 2176
Db 1396 KESSSG-----KREPSDK-----LRTAFIESAVERPFLMDRKKKEML 1434
QY 2177 ----- 2213
QY 1435 AOKKAELONGGLDLYSSGDGDPDN-----ONKLLNGVIPPDLFLAMEYTLGDYRDLIVHGN 1492
Db 2214 SMDKRTTVNSVTYILNENNRKODKKDELRKI-----FWEK 2253
Db 1493 TSDSGNTGNSNNNNIYLEASGNKEEDMKIOEKIEDILKNGSTPLVPKSSAQTPDKMNE 1552
QY 2254 NKKFTMEGIYLYT-----HLDNEKEKIRDN----- 2283
Db 1553 HAESLWKMICALIYTERKNPDTISARGDENKIEKDEVEYKEFGSTADKHGTASTPTGYK 1612
QY 2284 -OYN-DATKL-----TPSLEFYKRPOLFRMTEMAEENCKRKKOLKLE 2327
Db 1613 TOYDEKVKLEDTSGAKTPPSSASOTPLISDYLVRPPYFRYLEEGONGCNCKRKKRLKQIK 1672
QY 2328 AGC-----KEYECNG-----SNDKTOECACACTYONFLKWKTEYEROREK 2370
Db 1673 EECMDGSDKYSYSGDEOCDDRDTSNEVSADLEGRSCGNSCFYKWKIKRKKKEVDKQANA 1732
QY 2371 FKDKDKKKYDYSTEDIEKATCAHEYLMLKLELGKNDGSCMKPSSQLPKTTQOS 2430
Db 1733 YSKOK-----TKY-----EGS 1744
QY 2431 QSDANDPESLIDYVPEEFNKCCEPELSKKGSMHTKITEPKIIPMNCVEKAAV----- 2485
Db 1745 KGAGLND-----HNKPCYKLG-----TCTDAAPFLNRKLN 1775
QY 2486 -LSKEAEN-NDITL-----KEKIPLESYTERESKSNSTNNPCDPKRYADKTYIGRN 2539
Db 1776 GPCKKDNENGNDINGNTEEFERPAENCK-----PCSSFK-----INCRN 1816
QY 2540 -PCENNEENRFYDYEMKCYKNSKFTQEKRYCVPPRREHMLRLDEIKIERLDSNYL 2598
Db 1817 GNCRSGD-----DYKEKC----- 1830
QY 2599 LKAVRTANNEGIDILKNENSEGAMPICDITMYSFADIGDIRGDMIRIGYLPV 2658
Db 1831 ----- 1851
QY 2659 EIKIKVFEYITGKRRNKNKKNKYNDVOTFRSAMWANDRDIWAMTKAKAPEDAKLPK 2718
Db 1852 ----- 1898
QY 2719 GRMDF-----ERTLLDQCGHKDPPVDYIPQRMMTESEYCYKALMELEKFKSC 2775
Db 1899 EKGNAINDQITLYR-----ALIKRWEYF-----LEDYKIRKKKL 1935
QY 2776 DICKTSDRCKNDYDENKCEQCTRCQERYKNFYLKWKSILFDIOSNKYKELY----- 2825
Db 1936 KPC-----IENGNGSTCINGCNKKNRYGEMIKLKQEMWIKINHYLKKRKEGDN 1986
QY 2826 -----EOPLYTKISYVD-----HYQNFYQIKLT-----FKSCSVSESEYUHTSKCLNRYN 2874
Db 1987 VTSLATVLETLVTDIAANRREOTSLDKLTSLGNCBPNS-----R 2030
QY 2875 ENDGSSNIRTYAFETPRKSYEACSTLPKSNPLDNCPTDQNKDGKELQTFECSKNDY 2934
Db 2031 KNDGNE-----DAIDCM-----NKLEKIKHECTQH 2058
QY 2935 DNNIDNNMAYLYVNSSD-----NKGYLPPRRRHLCYRITAYNRKODKELKLLTSA 2991
Db 2059 ENSVEN-----SDQPHNCGSNPP----- 2078
QY 2992 FSOQLLGOKYKSEBELCFEAMKYSYADYSDIIGTDMDSLSSEKIKLFTSNENATEN 3051
Db 2079 -----DEEDLLER----- 2087
QY 3052 RKTWENNRROIWAMLGKRYATSKYTLDEGMCOLPKDETNQOFLWLEMAKQACKEK 3111
Db 2088 -----ENPVEQ-----PGFCPTQOQRPBD----- 2107
QY 3112 KHVSDSLTKCPRSNEDNFEASELLRQPGCONDIRKYSILNLIKNTMENLITKYQLKD 3171
Db 2108 -----DKCGKLEKKDEKE-----OP----- 2124
QY 3172 OSSGNIDNKPSEENVOSYIKSDSOCALINDINEITVGTKNENNEFKVLEKLYPGLY 3231
Db 2125 -----EOPAEEDGALVPSG-----PG----- 2142
QY 3232 FVEDETHKNVLDGNTKEEQGYRPAKALYFTPHVDSFOAPLPSTHRAVQAQDPKNDILK 3291
Db 2143 -SEPEADKGYVKAEL-----PKP-----QEPDLSHPA-----VIP 2173
QY 3292 SSISVIVTSLG-----IALHFKKKFKSSV-DLRLILNPOGEYMGPTLESKNRYIPR 3346
Db 2174 SLVTSILAMSVGIGFAFYFYLYLKKTSYVGNLFOILQPKSDYIPIFKLSPNRIIPYT 2233
QY 3347 SGPKYKTYIYMGDT-----SGDEKTYMMDLSSSDITSSSEYEELDINDIYVPGSKYKT 3403
Db 2234 SGYTRKRYIYLGDSGTDSGYTDHY-----SDLTSSSEYEEMDINDIYVPGSKYKT 2287
QY 3404 LIEVLEPESKRDIPSD-----DTPS-----NOTPRKRTIDEMNEL 3440
Db 2288 LIEVLEPESGNNTTASGNNTTASGNNTTASGNKTPBDQNDIQNGIIPSSKLTDEMNOL 2347
QY 3441 KHDVSOYL-----PNTPENNNYSADIPANTPENTLYSDPEKPEPLISIHDLDTGKEI 3497
Db 2348 KDEFISQYLQSEBNTPE-----NMLGYVNDNNTHTPTSH-HAVEKPTIMSHDRNLFSGEY 2404

QY 3498 SY-----NINMSTN-----TNNIDIPNARNDSTYRCIDLINDSLVTLMLTY--M 3539
Db 2405 NYDMFNSGNPNINISDSTNSMDSLTSNNHSPYNDKNDLYSGIDLINDALSGNHIDYDEM 2464
QY 3540 MK 3541
Db 2465 LK 2466

RESULT 7
T28161
hypothetical protein FC93-var11.1 - malaria parasite (Plasmodium falciparum) (fragments
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28161
R:Herandez-Rivas, R.; Mattei, D.; Sterkers, Y.; Peterson, D.S.; Wellem, T.E.; Scherf,
Mol. Cell. Biol. 17, 604-611, 1997
A:Title: Expressed var genes are found in Plasmodium falciparum subtelomeric regions.
Reference number: Z20483; MID:97154495; PMID:9001213
Accession: T28161
Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2647 <HER>
A:Cross-references: EMBL:067959; NID:91794255; PID:91809295; PIDN:AAC47438.1
A:Experimental source: strain FCQ27/PNG
C:Genetics:
A:Introns: 2158/3
A:Note: FC93-var11.1

Query Match 13.8%; Score 2686.5; DB 2; Length 2647;
Best Local Similarity 23.6%; Pred. No. 2.1e-117;
Matches 889; Conservative 379; Mismatches 861; Indels 1635; Gaps 120;

QY 31 SHSASNVLERYAKNIRHPSKYAKEHVS-----LKGLTKAEFRG-----PSTPVNK 79
Db 9 TNSAKAEVDEIGETIQ-----KKAHSDADTPRSQKNGNCEAKFYNGGIMQNSKL-- 61
QY 80 HNTYTPPCNLHKEHTNLRDYDNLRHPCGREDNREDESEEC-GNKIRNYKRRNDA 138
Db 62 -----CELDTIDTIDTVDGHSN--PCGQGVAFPPDNRSSQCKKRIKD-SVDNSV 109
QY 139 IACAPRRRRMCDKNLEALNDIN---TONHDLGNVLTAKREGSIVNNHP--HKGT 192
Db 110 GACAPRRRLHLCNLSIESIQTNNDSSKAKHNLAEVCYAAKEGESIVNYEQLGHTT 169
QY 193 SDACTALARSFADIGDIVRGIDMEKPNVHD-----KYETGLREYFKIH----- 236
Db 170 EGCCTALARSFADIGDIVRGIDMEKPNVHD-----KYETGLREYFKIH----- 236
QY 237 DGMEDVKNNDPDSGNGYKLRAMWNNRNKVEAITCDASYKSGYFMSQESNTPLFS 296
Db 230 NGAEERYK-----DGSNGYKLRAMWNNRNKVEAITCDASYKSGYFMSQESNTPLFS 296
QY 297 NPGCGHKGQVPTNLDYVPPYLRWPFDEMGEEFGKRIKAKYKDSR---NKEKRLYCS 353
Db 285 GEEGQCIDGVPNTLDYVPPYLRWPFDEMGEEFGKRIKAKYKDSR---NKEKRLYCS 353
QY 354 HNGHCDTITWKGGLIHLDKCTDCSTCKYFEVWLNQOEAFFKOKEREKEIOSYLSN 413
Db 345 RNDYDCTKTRISIDKXSMNECTKCLYCDPYKAKIDNKKKEFEKOKKEENELRYNNE 404
QY 414 DNFVNNINSEYKOFYEKLETOYATNDTFLNLNEGKCKGGLPEKDTITFNSADK 473
Db 405 SONSPKNYNNMYETDEYGNLKDQYOSND--FLKLINSEPTCTNIDAKSKIDFTK--DPE 461
QY 474 GIFYRSYCOVCPDGVK--CDGIKTYHKSDNREARNNDYKPRPVKPTNTIVLSCN 531
Db 462 ETSISHEYCDPCPCWCGKATGADG--TWKRLYENDPQCPKRYEPKVELETEDVLTGK 520
QY 532 EGGDITOKLENFNSSTNYKDNKNNQKWEYKNDENIRCKLEONTEINNDPKIISFHN 591
Db 521 ENNDIIVKLEAFCKTOSNTGFK--NEEMWCYGYGN--DKCYLNGEELGEGK--KVKQVDF 577

QY 592 FELMAYTLRLDTIKWMDKLTCTINNFTTTCIDECNRNCLCFDRWYKOEEMNSIKKLEPT 651
Db 578 IMFVVAHMLKDSIEMSKSLNCLSKDKTKCTICCNQCCQYDKWIGKVKHWTQIKKHF 637
QY 652 KKKNIQ-QSYSNINNNFEG-YEFKWMKDKDDEAKWIKELMENIKKRNKFSNLNNRDY 709
Db 638 KQTFQGWCHFYVLETVLEBDOFTTTRKAYGARELVHIOEMLOKKKEOV--LHEDASN 695
QY 710 LENAIELLDHLEKETATTC-----KDNV-----TNEACETSHNATTNPGVPRGTOPTK 759
Db 696 MKTILDELDELHEKAEKQCVLNHKNKDNCPADLSDSEEDIEDIQRONKCKAP--SGNHRA 754
QY 760 NIKELAOYFRSAVEEARNGL--HKLKKAHEGITYKGRKDKONLCRIMIKHSNRL 818
Db 755 LVNKAASNMMHKKRROLVNRGVSSTKLGDAAKEGRKSGTTIKLD--ICSTIDHSNAKR 813
QY 819 GFENSGCDGCTGDTGDTOTRRVVGTEWEVDEHNRKDHEDYIMPRRRHITSLNLEHOT 878
Db 814 GHVDQPCRRKDSKVYAKNR-----RWDYAGFTSNTYKDIYMPRRHOFCTSNLETLQTT 868
QY 879 DHPLENGIYD--DIVNSFGLDVLISAKYEANKIIRMYEKKNNLKGPKVTPDKHOTTT 935
Db 869 NKLINGNDINGNINIINDSLGDLVLRANTADPFIKMYKQN-----DYKNATY 919
QY 936 CRAIRSFADIGDILNGR-----DLERNGDVYKLGHLTFVGNHRSKLG 982
Db 920 CRAMKYSFADIGDILNGR-----DLERNGDVYKLGHLTFVGNHRSKLG 982
QY 983 KGNDRYNDAPKYLKLENNWENANRAKVAWAMCDIKYLDKSGHOSSTOSYCGSDHTP 1042
Db 972 -----KEDRP--YTKREDWENARRKIMWAMOC-----PPLNGSFPCSKSHIG 1014
QY 1043 LDDYIPQKLRMTAEWYCKVOKEEDKLEKCEKDKONGOGCTKESGCTCTKTEA 1102
Db 1015 LDDYIPQKLRMTAEWYCKVOKEEDKLEKCEKDKONGOGCTKESGCTCTKTEA 1102
QY 1103 CNEYNDITGL--WKQWNIISDKYKELHQAQMSVNSGIEASTAKNIDRNVEFLS 1159
Db 1070 CKNVYKIKNLLINGKEWDMETIKYKLYLOAOTTAANGRPDYSGLVDENERPVYVFLF 1129
QY 1160 ELYOONCK-----SNKSGTSDSAVIGNTYENYCAVLH--DTGNFDDCOQONF 1209
Db 1130 ELYKENGKIGNDPDPARRKSRRETAPASVANKDYSTAGVYHDMGPHMECKTQTER 1189
QY 1210 CDEKSDGKDNKAYAFDRKPDHAGACGCKSGSKPTVQIYTKK-----KAEKD 1258
Db 1190 C-EKTDQYVNTTFRKPPQYADACICNTRPPPKEDSRKRSDEEYKATYEBKA 1248
QY 1259 TE-----CTYNDILKENDGKQVBDCHPKKNSNGP--DMOC-- 1294
Db 1249 TEDAVDTGPPAPREKATTTLDVCPYAGVLT--KENLENACTYKGPAPKAPSMKICPT 1304
QY 1295 -----GNINLVDPDR-----VCMPPRRQKLCY--HFLANDN 1323
Db 1305 EKTNAATGSGSSGNGALQRAKRAYEESGPTVNSGSCITIPRRRLTYQKLHMAVSGN 1364
QY 1324 EIKKLOSQV-----NLKEAFIKSAAEFTFSYUYYSKSGKEGNEJDELKEL 1367
Db 1365 TVVSGAQTQOGGTSRSGKETPSDKLRTAFISAALETFFLDKRYK--KKAIAKKE 1420
QY 1368 KEKG-----IPPAFLSMFTYFGDYRDLFTGTD 1395
Db 1421 KKKOMVDSPLSTADPHNNVSLVIAPNPNYKTCVLPPLFLROMFTYLDGADYDIFGKN 1480
QY 1396 ---ISKGHE-----GSKLKEQIDSLFKNGDDKSPNGKTRQEMWTHSHIEAMICAL 1446
Db 1481 DIVIDTKNGDKDAIEREKTKIDAIERYLKNADQSPDERKQRTWEGNGHINMGICAL 1540
QY 1447 VKIGAKKDDFTENYGVNNVFSKSTYLEEFARPOFLRWLTWYDYCYTBOKYLLADVO 1506
Db 1541 T----- 1541

Qy	1507	EKKSDQJQJCTECCNKKCEDVYKMKKKKMMIIPQDYVYDEDEDKKRPDRHIGWATDY	1566	
Db	1542	-----YK-----	1543	
Qy	1567	TGYNATDYLNRRKFTASCGDQKPGSASVYORNIOTLERQAYYDADKHCGCTKTEINDKTYN	1626	
Db	1544	-----ENDEKCT-----	1550	
Qy	1627	ISSKDKCGLVKEANTGAIKMQNKGPNNYNKKETEDVLFPSRLRLICFHALDGNATYD	1666	
Db	1551	-----	1550	
Qy	1687	EYKDENGLRKLMEVAATEGYNLQYQYKKEKKEKIKTSPAHKSYEYPCSAMKYSFYD	1746	
Db	1551	PLKONGEGLKSALMD-----EKNNKKPKDQ-----KYQ-----	1577	
Qy	1747	LRLIIGIDNDEDEKOTEEENLKRIKFNKNTS-----YKGSOSTIGNPGSTAKRTFMNEN	1802	
Db	1578	-----DKVXLIDEN-----SCTSRKIYVPAKPKPTTTPPPSPSPF-----	1612	
Db	1803	KECVNNAMICGYKRGDRDGSNGSARSDEDLKCGSVSPDDYPMGKNNDETAQOFLRM	1862	
Db	1613	-----SRP-----PIRY	1620	
Qy	1863	FAEMGEDCKHKEKELEKELVAGACDITYCGNEDKRRKCT--DACTOYKRTISEMKPYOE	1919	
Db	1621	LEEMAEYTCREKRRKLEKTI-----KYECMDDEKOKCGSGDEDEEETIRK-----	1665	
Qy	1920	KOIKRTYGENKDKRIYSEHPVAKDAEDAREYLDKOLKICCNKSGDCEYCMQMDVSTQRLTD	1979	
Db	1666	-----QDYSTVYDFFVCPGCGKYC-----	1684	
Qy	1980	GNSQMPASLDDEKFEYEGKCNQVPRGPVYARRETPSPRSVLISKATASKREAKTAPPT	2039	
Db	1685	-----	1684	
Qy	2040	KOPKVENIATTEMBAQOTRTRRAAQOTRRKRTSTATTTESDVGVMYKALISNKPDSRGLEG	2099	
Db	1685	-----	1684	
Qy	2100	CNPKTYGQYPMKGCIVAKSKENENGICMPRRKRLCINNIOYLANETEKNRONDKEAFTI	2159	
Db	1685	-----FYKWM-----ICKKDEYD-----KORE-----ANNOKTARBN-----	1715	
Qy	2160	KCAAIETOFLMKLIENPAENELONGTIPDEFKRIYTYGYOVKMDFFETDISDKTI	2219	
Db	1716	-----NDNAF-----	1720	
Db	2220	ITVYNSVTTLIN-----ENNNKQDKR KDE--ELKRIFEWKNKKFELWBMITYGLYHLTDE	2273	
Db	1721	STTDTCTGTAGDELQTLKNGPCKNNDVDSGENKKLIF-----	DE	1759
Qy	2274	NEKEKIDNTQYVDMTKRLPSPLEEFYKRRQFLRMFTBWAEEFCMKRREOLKLEAGCKEY	2333	
Db	1760	N-----GDTIKTY-----OYCG-----TCSLNG-----	1778	
Qy	2334	ECNENSDNGKTOEACAECVYONEFIKMKWTEYEROREKFRKDKDGKYYKDYPTERDIEKA	2393	
Db	1779	KCNQGD-----CRVRYTN-----	1791	
Qy	2394	TCAEYIYNNMLKJELCNCNKCSCMOKPSSOLPPTTQOOSOSDANMPESLDYVPEEFNKCE	2453	
Db	1792	TC-----NGSNRRTTITADDI-----	1807	
Qy	2454	CPELSKSGMIHTKRIITEPKIIPMNCVEKAAAYLSKEAMNNN-----DITLKEFPIESTRK	2509	
Db	1808	-----KNG-----GNSAFINLWSDDINSNGFNDELA-----	1835	
Qy	2510	EKESKSNWNNPDCPKKYYADPKYIGRRNDCENRENRK--VDYEMWCKYNSKPYQEK	2567	
Db	1836	-----CKN-----ANIFGJIKENMKK-----YFCK	1857	
Qy	2568	KRVCPVPRRHHCLRLNDELKIERLKDSNYLLKMWVRTANNEGDIILKFNNSNGCAMP	2627	

[illegible]

RESULT 8
T28157
erythrocyte membrane protein 1 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: T28157
 R:Yang, S.O.; Kollisch, W.S.; Gut, J.; Wu, J.; Ahn, J.; Petersen, C.; Fujioaka, H.; Aikawa
 submitted to the EMBL Data Library, July 1995
 A:Description: The molecular cloning and DNA sequence analysis of Plasmodium falciparum
 A:Reference number: 220479
 A:Accession: T28157
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2212 <YAN>
 A:Cross-References: EMBL:U31083; NID:g1517813; PID:g1517814; PIDN:AA06961.1
 C:Genetics:

Query Match 13.8%; Score 2677; DB 2; Length 2212;
 Best Local Similarity 31.1%; Pred. No. 4.5e-117;
 Matches 713; Conservative 307; Mismatches 800; Indels 476; Gaps 83;

31 SHRSANVLERYAKNIRHPSKYAKEHVS-----LKGDITKAEPFG-----PSRPVKN 79
 9 TNSAKEVLEDEIGETIQ-----KKAHSDADTFPSQLKNGEAKFYNGEIMOPNSKL-- 61
 80 HNYEYPCNLDKHEHTNLRYPDVLNRHPCGHEONRFDEDESEC-GNKTIRYKRNDA 138
 62 -----CELDHTIDTNTVDGHSN---PCEGQIVAFRPDNRNSQCKNRKID-SVDNSV 109
 139 IACAPRRRRHCKNLEALNDIN---TONIHDLGVLTAKTEGESYANNHP---HKGT 132
 110 GACAPRRRLHCKSHNESIOTNTNYSKSKHNHIAEVCYAKFEGESIVANQEOLGHTT 169
 193 SDACATLASEFADIGDIVRGIDMEKFNVDH-----KVETGLREVFKKH----- 236
 170 EGCITLALASFADIGDILIRKDKYLGNPQESARKOLENLRKIDFTYIELTSSRNKGT 229
 237 DGMEDVKNYDNDGSGNYYKLREAMNNVNRNFWWEALITCDASYKSGYFMOSESNTPLFS 296
 230 NGAEHYK-----DGSNGYTKLREDWMNARLIDIKWAMICKAPGNAQYFNTOSNGEKP 284
 297 NPGCGKQKQVPTNLYVQYLLWFMDEMGEEFCKRRIKAKYKDCR---NDKERLYCS 353
 285 GEGKQCIDGTPNPILNLYVQYLLWFMDEMGEEFCKRRIKAKYKDCR---NDKERLYCS 344
 354 HNGDCTTITWKGILHDKRCDTSCCKVFVWILGNOOAFKQKQEKYKESIOSYLSN 413
 345 RNDYDCTKTRISDYSANRECKTCLYCDPYKWKIDNKKKEFKQKKEENIYRNES 404
 414 DNKFVANNISEYYKQYKELKETQYATNDTFLMLNGKXKGGLPENDITFTNSADK 473
 405 SQSPKRYNNMYETDFGNLKDQYQSMND-FLKLNSFETCTNIIDEKRIIDPTK--DPE 461
 474 GIFYREYQOVCPDGVK--CDGIKTHKSDNDRKRYNNEDYRPMGVKFTNTITVLSGN 531
 462 EYFSHREYCDPCWCGKQKQADG-TWKRILENDPOCEIKRYPBPQGEETEDIVLTGK 530
 532 EGDITQKLENFCSSTNYKDNKNNQKWEYKNDENIRKCLQONTLEINNDRPKIISFHN 591
 521 ENNDIYLKREFCKTGNGFK-NEEMNCCYQYGN-DKCYLNGESELGGRK-KYKDYDNF 577
 592 FELWTVYLLDRTIKWMDKLTCTINNNTTTCIDECNRNCLCFDMVAKQEEBMSIKLFT 651
 578 LMFVVAHMLKDSIEMRSKLSNCLSKDKCTICKKNDCKCYEKWIGKQKVEYQIKKHF 637
 652 KKKNIQ-QSYYSNINNLFEQ-YEFKWMKDKDREAKKELMENIKRKNFSLNNRDY 709
 638 KQDFQGWGMYFVLEVLBDQFDTITTKAYGAREIVHIOEMQKKREOV--LHEASN 695
 710 LENAIELLDHLEKATATC-----KDNNTNEACETSHN-----ATTYPCVPRGGTQPTK 759
 696 MKTIIDELDHLEKAKQCVNHNKDNKCPDSDSEDEEDIDIQQRNCPKAPGSSAARAL 755
 760 NKEIAQYEFKRSAYEEARNGCL-HKLKGAHEGIVYKRGGRKDKONLCRIMIKHSNRL 818
 756 -VAKVSNMHHKRRRLQVNRGVGSSKLKGAKEGYRSGTITIKLD-ICISITDHSNAKR 813

819 GFSNGPCDGTGDTGDIOTRFRVGTWEVDEHMKRKHEDVIMPPRRRHITSNLEHLOT 878
 814 GHDDQCKRRK---DSKSEMFRTEDGCK-PAGFISKTYKDYMPBRORHOFCTSNLEYLOT 869
 879 DHPNGNIVT---DLVNSFLGDLVLSAKYEANKIIMYKKNLKKPKRVTQPKQGTI 935
 870 NKLINGNDIGNNRIINDSLGDLVLAANYEADFIKMYKKON-----DYKNATI 920
 936 CRAIRYSFADIGIIRGRDLMERNGDMVAKQHLLEVFNHSLKKGNDKYNDAPKY 995
 921 CRAMKYSFADIGIKOTDWMNDGSESKTRDLRELFTDIKKKHPI-KEYKEDPY-X 978
 996 LKIRENMENAKWEMAMCDIKYLKDKSGHOSTOSYCYGSDHPPLDYIPQKLRMT 1055
 979 TKLREDMWENARRKIWEAMOC-----PTPNPSFCKSHHTPLDLYIPQRLRMT 1027
 1056 EMAEWCKYQKKEVDKLEKCKECKDKDNQGGCKRESGCTCTEACNENYDIIQMK 1115
 1028 EMAEWCKEOKKOYGEVLVSCNGCKDE--GVCITNKS-SQCTSCMQCENYKNEINTWKE 1084
 1116 QWNIISDKYKELHQAQMSVNSGIEASSTAKNHIDNVEFLSELYOQNGK----- 1168
 1085 QMDKMEIKYKLYLQAGTAAAGGPDYSGLVDENEKPVVNFLELYKENGKIGNPDRT 1144
 1169 --SNKSGTSDSESAVIGTNTIYENVAYLH-DTGNFDCQSONEFCEKSPKQNEKAFR 1225
 1145 PAKRSKRRETPAPSAVANDYSTAAGYVHQBMPHMECKQTEFC-EKTEQVYENYTFK 1203
 1226 DKPODHDGACGCKSGSKPTVOIKTK-----KAEKQTE----- 1260
 1204 NPPQYDADICITNTPPRPKDSKRSDEDEKVKTEKATEADAVDTGPPAPKEA 1263
 1261 -----CKTVNDILKENDGKQYEDCHPKKNSNGP-DWOC-----GNT 1297
 1264 TTLLDVCPIYAGVLT-----KENLENACPTKYGPAPTSWKCIPTEKTNATVSGSSGNG 1319
 1298 NLYEDPR-----VCMPPRQKLV---HFLANDNEIKLDSQV----- 1332
 1320 ALQRAKATVESGSPVTSNGSICIPRRRRRLYQKLMDASGTVVSGAQTPOGTS 1379
 1333 -----NLEKAFIKSAAEFTFSWYKSKDGEENEDKELKE----- 1369
 1380 PSKRETPSDLRKAFIOSAIEFTFLMDRYKE-----KELEKKEKAVANGVPSLNGP 1435
 1370 -----GKIPAFILSMFYTGEDYRDLFGTD---ISKGHE----- 1402
 1436 PQCPGYTGDSPOSLQGTGVIIPPELFQMFYTLGADYADIFFGKNDIYIDRKNQKDAER 1495
 1403 GSKLEKIDSLFKNGQOKSNGKTRQEWTEHSHETIEMALCALV-----KI 1449
 1496 EKKIKDAIEVLNNAOSQPSDEKROTWMQNGEBHWNMICALTYKEKDEKGTPLKONE 1555
 1450 GAK-----KDFETENYGVNNVYKFSKDS-----TTLSEFAKRPQPLRLTE 1489
 1556 GLSALMDENKKNPKQOKYQYDKAKLDENSGTSTKTDHPPPLTNTFISRPFRLEE 1615
 1490 WYDYCYTRQKYLKDVQEKSSND--OLKCD-----TECNKK 1525
 1526 EDYVYMKKKK-EMIPQDKYKQERDOKKRRDRQHIGVWYDYGNTATDILNKFATSC- 1583
 1676 RPYKRWLEKKRDEYDKQEKYNNOKTDARRNNNDNASTTLDCTCTTAGDPLQTLKNGPK 1735
 1584 -----GDKPGSASVQNRNIOLEKQAYYDADAKHGCCTFIENDK-----YTNISSKDKGKL 1636
 1736 NDANVDSGKKNKIFDEGDFKTYQYGT---CSLNFCKNGDGCRRRTIVT-----CNGS 1788
 1637 VKEANTGAIKQWNGPNNYNNLKELTEDVLPPSRRLICFHALDGNATYDPEVADENGLK 1696
 1789 NRTTITADDIKNGG--NSAEINNLVSD-----DINGNGFN- 1823
 1697 RIMEVATIEGVNLGOYYKKEKKEKIKITSDAHKSYEVPSCAMKSYFDLRLIL----- 1752

```

Db 1824 ---DLECKNANIKFKIKENMK-----CYFECKSDYGLKKNNDIDONQIILRAL 1872
QY 1753 ---GIDNLEDEKOTENKIKFKNGSVGKSGDSTTGNGSARFKFFNENKECWMA 1809
Db 1873 FKRLLEFLDYNNKIRKKNPCIN-----GEKAITNGCYEO-----WINORTEMT- 1920
QY 1810 MICGYKRRGDDGNSGNSARSDEDIKCGSVSPDDYPGKNNRDEGTAYOFLRMEAGEMD 1869
Db 1921 -----NIKN--RNEOY-----NSGDDTEM--KSPFSLVYLILROA---AT 1955
QY 1870 FCKHKEKELELYGA--CNDYTGDNEDNRKCKTDACTQYKFKPSEKPOYEKOIKYGE 1927
Db 1956 IDKGNHNGGLKLYKSVACN---CGNSONGKE-----GE 1986
QY 1928 NKDKITSEHNVANDADAREYLDKOLKICEN--KSGDCEYKCMKDYSTORLMD----- 1979
Db 1987 ENLDVYL-----LLOKLEKKAKEKCKDNBPETSGIPQPC--EVSFPHIDEQPLE 2034
QY 1980 --GNSOMNPASLDD---EKEVEGKNCQVPPGPRVRRPETSPPVSLSKATASKKEA 2033
Db 2035 EENVTJEHPKICDVLKTOPQPEPGBETCBESGPTDYKEEEREEK-----KEEKDEE 2089
QY 2034 KTAPTKOPKVENLT 2049
Db 2090 PGLPPTPPAPSPST 2105

```

RESULT 9

```

T14602
variant-specific surface protein - malaria parasite (Plasmodium falciparum) (fragments)
C:Species: Plasmodium falciparum
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
C:Accession: T14602
R:Yoss, T.S.; Felger, I.; Weiss, N.; Beck, H.P.
submitted to the EMBL Data Library, February 1998
A:Description: Identification of a conserved 5' flanking region of Plasmodium falciparum
A:Reference number: Z18158
A:Accession: T14602
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2135 <VOS>
A:Cross-references: EMBL:AF050740; NID:g2944094; PID:g2944095; PIDN:AA05220.1
A:Genetics:
A:Gene: varphi17

```

```

Query Match 11.9%; Score 2303; DB 2; Length 2135;
Best Local Similarity 29.4%; Pred. No. 1.4e-99;
Matches 681; Conservative 306; Mismatches 745; Indels 584; Gaps 95;

```

```

21 GDAKSPIIKESHKSARNVLERYAKNI--RHPSKYAKEHYDSLKGLDTKAFPGGSPSTPVNK 79
Db 5 GPOGGGTRKD--DAKHMFDRIGTIGQOVNNAAEFFNEILOGRSSA-----TNTNL 55
QY 80 HNYVYFYPCNDHK--ERTNLRTDYVNLRHPCGHEONRDEDEESE--CGNKRIRYKRN 136
Db 56 KLIGNIETCSLEKKYKHPNGGVSDBKRYPC--KELSRKYVDRPSDKIGGCTNEKMR 113
QY 137 DAL--ACAPRRRHMDKMLAL--NDINTONI--HDLGNVLVTKYBESEST---VNNHP 189
Db 114 DGIAGCAFYRLHLCHINLETETENNYSNNAKHNLDVCMARKYEDSITKYTYTGQH 173
QY 190 K-GTSDACTALARSFADIGDIVRGIDMKFPVNDK-----VETGLREFFKTIHMGDEY 243
Db 174 KYDSOLCTVLAARSFADIGDIVRGIDLVGNTQKREKREDLENTKDLFKTIHGLGCV 233
QY 244 KNDNPOSSGGYKTLREANVANKYWEATTCASIKSGYEM-----QSESNTPLFSNP 298
Db 234 KARYK--DTNTEYELRDEWMTANRETYKAITCDA--HGTYFRATCGGDNSPSMAKNNC 289
QY 299 KCGHROGK---VPTNLDYVPQYLRFWFDWGEERCRKRNIKLKVKVQSCR--NDKRELY 351
Db 290 RCGKKDGRPDQVPTLYFDYVQFLRFWEENAEDECRILKRLKDALOICRKEHKGARKLY 349

```

```

QY 352 CSNHGDCSTTTIKKIGLILDNKCTDSTKCYKEVWMLGNQOEAFFKQKEKEIOTYL 411
Db 350 CDLNRDYCEOTASGKHDFEEDCKGCOYSCARFVWIDNOKKEFLKQKKKGTLE-SLK 408
QY 412 SMDNRVYNNINSXY-----KQYEKLEKFOYATNTFLMLBEGKYC---KGLPGEKDI 464
Db 409 SRKRKAGISITKYVDGYRKEFEKLEK--SEYRVGEFFLILNNETCKEVEKDKKEGKIDF 467
QY 465 TETNSADDKG---TFRSEYCOVCPDGVKDCGDKTYKHSKNDNRERNNNDPRPMGV 519
Db 468 KTVNSGASAGDVNKTFFYTKYCEACPWGAGDERNGVMAKADRDGSPGDYK---Y 524
QY 520 KPTNIVYLSNGEGDITQKLENF-----NSSTNYKD 552
Db 525 KKEEPIILGDTKSEIYERKFKCKNKGANGREGVGSGEANGAASNSDNATTYCG 584
QY 553 KNN-----QKWECCY--KDN-----INRCKLEONTLEINNDPKTISFHFEEELMYTL 599
Db 585 GGNDSLSLEKWTQYKKKKENNDGKDINFC--VQAMONSKKDKQVASTNAFMDWHDH 643
QY 600 LBDITKMDKLKTCIN--NTTHCIDECNRNCLCPDRWVKOREEEMNSIKKLTFTKKNIQ 657
Db 644 LIDSIKWRNEHGKCIKNKGKT--CIKCKSKDCFLKVVQKEKEMKILHEHNTQGGFD 702
QY 658 OSYNSINNNLFEGYFFKYMDDKDEAKKEMIKKKKESMLNNRDLF----- 711
Db 703 KGEHRLGFTHDVLYNLTLDK-----KELKTI--EGTYGNTERTKIKKELDKEE 751
QY 712 -----NATELLDLHLEATATICKDN-----TNEACEESH 741
Db 752 KDAGGTGVASGTPKRNIMDKLIEHLEDEAKKCKOQEPQSLGSLNPHVYDDGSPKR 811
QY 742 NATNPCKVPRGGQPTNKEIAQYF-----KRSAYEARNRGLHK--LKGAEHGIYKR 795
Db 812 DRTNPGCSDT--TEYAVLAKVAKQKQGEVRAKMLERSRKNKTSLEEDIKKAQFKN 870
QY 796 GGRKDFDNLCLRIKHSNRLG--FSNGPCDGGGTGIGITRVYVTEWVDEHNRKD 854
Db 871 GRSSELNGDICKIDNKXSNDRSTAGPCTGK--DGGERRNAOTKMGWD--NFSAT 926
QY 855 HEDVIMPPRRHHTSINLEHQTDDHPLNGNIYDVLVNSFLGDVLLSAYEANKIIRMY 914
Db 927 HKNLYIPRRQHMCTSNLEKIDFLSVTSKSN-----VNSDFLDVLLAANNEAQRTHDF 981
QY 915 KKNNLKGPKEVTPDKKHQTTICARIRYSPADIGIIRGRLMENGSMVLOGLHEVFG 974
Db 982 AHKKDDHG-----IACRSVRSFADLADILIRGRMDKDDGAQKMEIDFKFIKG 1030
QY 975 NIHSKLG--KNDKYND--APRYLKLRENWWEANRAKYWEAMKCDIKYLRKSGHOST 1030
Db 1031 NLVPSLFGING--KYDGDODRTPQYKQLRBDWWEANDQYWKAMVC-----EKDGKCD 1082
QY 1031 QSSYCGYSDHPLDDYIPOLIRMTWTEWAEWYCYQKKEYKLEKCECKDKONGGCTK 1090
Db 1083 E-----DTPVDDYIPOLIRMTWTEWAEWYCYQVQSEYDELKTKKSGSCKIKKVGCT- 1134
QY 1091 ESGTGCTKCTEACNEYNIDIGLWKEQNIISDYKXELHEQAOQSVNSGIEASTAKNHI 1150
Db 1135 SGSDSDCTPCAECTTYGQKIKRPWEDQNNMLLOYTILYWAETTAYAGGIRAYSGVQK 1194
QY 1151 DRNVIELSLYOONGSKSKSGTSDBSAVIGTNTYENAVGLHDTGNFDDQSOQNEFC 1210
Db 1195 DKPYOGLFLELOKONGSK-----TYNNPAAIYIHQEARVGECEVOYKFC 1238
QY 1211 DEKSDGKDNKCYAFRDKPODHGAGCGK-----SGSK-----PTVQITTKKAE 1256
Db 1239 NTNG-----NODKYFRKPKDHDACCKDKPOQASAGGAGANLPSRPVYDDDDHSSD 1295
QY 1257 KQTE----- 1265
Db 1296 EDEEEDDEDGDAEDENDEPASEEYKDDTEDEVVESTAVSOPAAPTTTTGGVTPACEIYK 1355

```

1266 DLK--ENDGKQVDCCHK-KNSGYPDMQC-----GNTLVDPDPR---- 1304
1356 DLEKRNKTFK---FACTORYGNNRSLKMKCIPTSGDPAATRGSGDPTTKQNDSEKSGSE 1412
1305 -----VCMPPRROKLCVHFLANDNEIKLOSQV--- 1332
1413 EGSORAKKHTSDASEKSAKSGEPGSGSICLPPRRKRLVGGITVMAELQSSQSLG 1472
1333 -----NLKAEIKSAAETFFSWYYK-----SKDEGNEL----- 1363
1473 OTTAPGTSOAOPLAARVESAAYETFFLMDRYKTLNAPQSGSSLGGAFLQILMGALG 1532
1364 -----DKEKKEKIPPAFLRSMPTTGDYRDLFGF--DISK-----GHGEGSK----- 1406
1533 SEETPTSLKSGTIPDFLRLMEYTGIDYGVKQVYKALASGDKSSKNPMQET 1592
1407 KEQIDSLFKNGDQKSP-----NGKTROEMWTEHSHETWEAMLCALY--KIGAKK--- 1453
1593 SKTIEILKNGCTPPPTPTPTPHSSGTTSSMKKTNGOHIMKMICALITYKESGKKIEQ 1652
1454 ---DDEFTENYNNVNFSD-----KSTLEEFAPKRPQFLRML 1487
1653 VKATDNTDLFEKLDKDSYDYKVEIKENETEAQGPDDGLTPQTLTSLNFKRPYFRYL 1712
1488 TETDYDCTTRQKYLKDVQCKSNDQKCD-----TECNK-----CEDYK 1530
1713 EEWGONFCRTKRMKLDIYECHNSDNGHDYCSGDGYCHNNELKHNHGHIDCQCYK 1772
1531 YMKKKKEMIPDKYKDERDKRFRDROHIGVWTD--YTG---TNAVDYLRKFTASCGD 1585
1773 ECKKRYKRW-HKKFDFHOKKNGYGEHEKTLNGWYSGGDNTNCCQKKEKKA--- 1827
1586 KPGSASVONRQTLKQAYYADKHCQCTKFLENDKYTNISSKDK----- 1632
1828 -----EDFLKALRHC-----KNSDDTDKSEDEKKNKINDKPEENTFNS 1867
1633 ---CK-----GLVKEANTGAIKMONKGPNNYNNLK-ELTE-DVLPFSRRRLICHALDGN 1683
1868 TYCKACPIYGVY--CNRGRGCRPKIINEKNNEEGEOTDINILID-----DGA 1914
1684 TPEVVDENGILRRKLEVAATEGYNIGYKKEKKEKIKTSDAHK-YSEVPSCSMKY 1742
1915 ND---TDMELHKKCKE-----YGLTYNLKQEMKO-----HNKSKCELOLRPLNS 1958
1743 SFYDLR-----DIIIGIDNLEDEKORTEENLKIKFNKNGTSVYKSGSDSTGNP 1790
1959 EYDEKIPKILFERWILFI-----QYNNKSKERITRCTN-----DVNSCKEGCKGNC 2007
1791 GSTARKEFFNENKECVNMAMICGKRGRODNGSGNSARSDEDLKKGVSDDYPMGKN 1850
2008 DCVEE---WLKKSAMEKIKIDYNOHFOVDVERIASRKSFEQ--GTFPS----- 2054
1851 RDEGTAYQFLRMPAEWGEDFCFKHEKLEKLVACNDYTCGDNEDRRKKTDACTQYKFF 1910
2055 -----YAKKAOEYVECKE--EOKELMGCTGSGNDGCTEE---CENG-----DF 2092
1911 ISEMKPOYEKQIK--KYGENKDKIYSEHPYAKDAED 1944
2093 ITNIDLRLQTKIQYCOTQHNKTQRNSCHPPPNDEE 2128

RESULT 10
B71600
variant-specific surface protein 1 homolog PFB1055c - malaria parasite (Plasmodium falciparum)
N:Alternate names: erythrocyte membrane blinding protein 1 (EMP1)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: B71600
R:Gardner, M.J.; Tetteh, H.; Carnucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Petter, M.; Salberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: B71600
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2197 <Gap>
A:Cross-references: GB:AE001434; GB:AE001362; MID:g3845341; PIDN:AAC71996.1; PID:g384
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB1055c

Query Match 9.0%; Score 1749; DB 2; Length 2197;
Best local similarity 20.4%; Pred. No. 1.2e-73;
Matches 751; Conservative 323; Mismatches 732; Indels 1884; Gaps 120;

33 KSANVLERAKNIRHPSK-YAKEHVDLSKGLDTKAEFRGSPSTPVNKNHYYPYPCND 91
12 ESWKHEDRIGEDYBQVKESEVYVYSELGKISLAPILGVEGSGTNE-----TCNLV 64
92 HKHTNLARDVDVLRHPCG---REONRDEDEBESCGKTI--RNYKKN--DAIACA 142
65 QDYNNPYVGNNSN-RYPCKNLKGITNEERPSDLGGQCTKKIKGNEYSTRSGKDCGACA 123
143 PPRRRHCKENLEALNDINTONIHLLGNVLYAKYEGESIVNNHP-HKGT-----SDAC 196
124 PYRRLHCSHNLSTID--TTSMTKLLLEYCAAKTEGNSIDTHYHOHQTNEDESPQIC 181
197 TALARFADIGDIVRGIDMEKPN---VHDKVETGLREFKKIHDGMEDE--VKNDYNP 249
182 TMLARFADIGDIVRGIDLEFGNSKEKEKDELETNLKTFGIHLEKNOKEGAEFRYGS 241
250 DSGNAYYKLEAMANNVNRKVPALITCDASYKSGYPMQS---BSNTPLSNPPC----- 300
242 D-TTNYOJLEDMWYANRAVWALITCDV-HGSDYFQTCGTGDETTATRYKDCRCKDEN 299
301 GHKOG---KVPTNLQVPOYLRFMFDEMGEPFCRKRNIKKYKDCSRNDKERLYCSHNG 356
300 GKRPGSNADQVPTFPDYVPOYLRFMFEMADFPKRRKKKIKLEKQOCROVKQMLYCSGNG 359
357 HDCTTIWKKGLHLNDKCTDCTCKVFEEVWLGNOGAERKQEKYEKIQSYLS--- 412
360 YDCTKTYKKGLVIGEHCTGNSVWCRLEYBSWIDNQLFKOKOKYETEISNGSGGGS 419
413 ---NDKFFVNNISE-----YKQFYEKLETOYATNTDTLNLNBEKYCKGGLPGE 461
420 GVGKRRKRRKGAGVEFATNYDGERKFEYKELSESGKVDLKLNLNNDYKCKIKRDEK 479
462 KOITFTNSAD---DKGIFRSEYCOVCPDGCYKQIKYTHRSNDNREVNNEDEKPPW 517
480 EKIDFTKPADKNNNEGTFFHSEYCKPCPGCYKRRDNOKKIDYDKCTR--GKIYEPAS 537
518 GVRPTNITVLYSGNEGODITQKLENFCN---SSTN-----YKDKNNOK---W 558
538 GAGGTPTKIKLSEKQEIETKLKAFCDQNGDTJNSVARGADGSGSKSKELEYEEM 597
559 ECVYK-----DENIN-----CKLE-----QNTIENDPKIISFH 569
598 KCYNEQVQYKDKNGEEDDEEDVDYKVRKAGLCLILENKHESRNNSSNEBPQOKTFH 657
590 NFEELWVTVLLRPTIKWMDLKTICINN-TTHGIDECNRCLEFDRVAVKQKEEMNSIRK 648
658 DFFYFMIGRLANDSMYRGVNSCINNPKRKKCRNECKDKDCGFKWIKGKKKEWEIKK 717
649 LFTKKKNIDQSYSNINNLEEGYFFKVMKLDKDEAKWKEIMENIKRKNNEFSLNNRD 708
718 HF-----KTQAFKN--KRENGSIMGSLMD 742
709 YLENAIEILLDLHKETATTCCKDNNTNACTSHNATNTPCVYKRGCTQPKNKIKETAYF 768
743 SAQVVELAL-----ELEQLP 758
769 KRSAYEARNRGLHKLKKAHEGITYRGRGRKDFKDNLCIMIKHSNRLGFSNGPDCGK 828
759 -----QIKD----- 763

QY 829 GTGDIOTRIVGTWEVDEPHMKRDEHDVIMPRRRHICTSNLEHLOTDPHNGNIYD 888
 Db 764 GYGD---VKELKJIKELDEEKKKQAEAAV-----V- 794
 QY 889 DLVNNSLGVLISAKYEANKIIMYKEKNNLKPKREYTDPKHOTTCIRAIRSFADIG 948
 Db 795 -----VAONOKKTTIDKLOHE----- 811
 QY 949 IIRGRDIMERNGMVKLOHLETFVGNIHSLKGNKNDKAPKYLKRENNWEANRA 1008
 Db 812 -----GDDANNCLKH----- 822
 QY 1009 KWEAMKCDIKYLDKSGHOSSTOSYCYSDHPDLDYIPQKLMTEMAEWYCKYQKE 1068
 Db 823 ----- 822
 QY 1069 YDKLEKCEKCKDKD---NOCGCTKESGCTCTEACNEYNIDIGKREOMNIIISPKY 1125
 Db 823 ---KECCEOTOPKPPAGPAPSETGETT-----LEDEE 856
 QY 1126 ELHQAOMSVNSGIEASSTAKNHIDRNVIEFLSELYOQNGKSNKSGTDESAGVIGTW 1185
 Db 857 EEBEE---EDAGDEV-----EBGETVDT-- 876
 QY 1186 TYENVAGYLHDITGNFDCCOSQNEFCDEKSDGKDNEXYAFRDKFQDHDGACGCKSGSKPTR 1245
 Db 877 -----TEGDETFVEOPVKDTPREG----- 896
 QY 1246 VOIKTKKAEKDETE---CTVNDILKENDGKQOVEDCHPRKNSG---YPMOC---GN 1296
 Db 897 -EBEEKAKATDTTSLDVCDTYKNAALNND- NLTDACKLKTGPGKEKFPNMKCVSSE 953
 QY 1297 INLV-----EDRVCPFRPRKLYHFLANDNEIKLOSQVNLKAPFKSAA 1344
 Db 954 KSVATPGSSGAGCKSGDKAGICVPPRRRLYYGGLK-----LTSACT 996
 QY 1345 ETEFFSWYKSKDGESENLKELKEGKIPPAFLRSMFTYFGDYRDLFTGDSIKSGES 1404
 Db 997 SS-----ESPQG-GSESSR-----ASDSQSGN--- 1018
 QY 1405 KLEQOISLFFKNGDQKSPNGKTRQEWTEHSHIWEAMCALVKIGAKKDPTEYNGYN 1464
 Db 1019 ----- 1018
 QY 1465 VKEFSKSTLEEFAPKPOFLRWLTEMVYDCTYRQKYLKDVOEKSKNDQKCTECCNK 1524
 Db 1019 ---GDDITTESLRK-----WFI-----TAA 1037
 QY 1525 CEDYKYMKKKKEWIPQDKYYDEDRKRFROHIGVMTDYTGNTADYLNRRKFTASC 1584
 Db 1038 IETFLMHRKKEWBAOKK----- 1056
 QY 1585 DKPGASVVOHIOLEQOAYYDADKHGCKRFTENDKYNISCKDKGLWEANTGA 1644
 Db 1057 ---AELORNGLL-----GTA 1070
 QY 1645 IKWONKGNNTNNLKELTEDYLFPSRLRIGCFHALDGYNTDPEYK---DENGKRLRME 1700
 Db 1071 S--JLGGDDSNPQTOLOKSGIPLDLRLMFTYJ--GYRIILYRGVADDDNG-- 1120
 QY 1701 VAAETGYVLGYYKEKKEKIKISDAHKIYEVPPCSAMKYSFYDLNIDITGNTDEDE 1760
 Db 1121 ---GN-----ITLNASGNKDE 1134
 QY 1761 KOTE---ENLKTIFFNKNGTSVSGSDSTGNGPSTARKFFWMNENKCEYNNAMICG--YK 1815
 Db 1135 KOKKEKIEKIEQILFTPSGNKETRGPONSVD-----ROSLMDRIAHEHVMHGMCAJLYK 1189
 QY 1816 RGRDGSNGSARSDEDLKCKGVSDDDYPMGNRDEGTAYOFLRWFALWGEDFCHE 1875
 Db 1190 ---DDDN----- 1194
 QY 1876 KLEKIVGACNDYTCGDNEDKRRKCTDCTOYKFFISEMFPQYEKOIKKYGENKDKIYSE 1935

Db 1195 ----- 1194
 QY 1936 HPVAKDAEDAREYLDQOLKICENKSGCEYOMKOVSTQRLTDGNSQMPASLDERE 1995
 Db 1195 -----LKV----- 1198
 QY 1996 VBEKCMQVPRGPVRRETPSPVSLISKATASKEAKTAPTRQPKVENITTEMBQ 2055
 Db 1199 -----VAKPKIEN----- 1207
 QY 2056 TTRBRAOQTBRKRTSTATTTESDVGTMKAILSNKPDSSRGIEGCPKTYGOYPRWGCI 2115
 Db 1208 ----- 1207
 QY 2116 GSKENENGICMPPRKIKLINNIQIYLVETENKRDNDIKFAIKCAIETQFLMKYII 2175
 Db 1208 ---PEK-----LM-- 1212
 QY 2176 ENPAENELQNGTIDEFKRLMYTYGYDKMFFGTDISNDKIIITYNSVTTLNENK 2235
 Db 1213 -----NETTK 1217
 QY 2236 KRODKDEBELKIPWEKKNKFIWEGMUYGLYHILTDENEKERTINDYQYMDTKLPSL 2295
 Db 1218 KPKDEKYQYQAKL-----EDESSEKRPDSASG--TKLT-- 1250
 QY 2296 BEFYRPOFLMFTWAEFCNKRREOLKLEAGCKEYECNSDGTQOECF-- 2348
 Db 1251 -DFIKRPPFYRLBEMGENEFCFKRTMELGKIKEDC- YKNGRSGSGGLKONEIVIDREK 1307
 QY 2349 ---ACTYONFIKKMTEYERQREKFKKNDGKTYIDYFSTERDILEKATCAHEYL 2400
 Db 1308 IFGDLCTCTCARHCFYKKW-----INTKRD----- 1333
 QY 2401 NMLKELCGNDCSCOMKPPSSQLPRTQOSOSDANDMPESLDVYPRFKCEPBLSK 2460
 Db 1334 -----EFNK----- 1337
 QY 2461 GSMHTKKITEPKIPMNCVEKAAAYLSKEANNMDITLKEKFIPIESTKREKSNWTNN 2520
 Db 1338 -----OSNAYSQKKRY-----EENDSAQKN 1360
 QY 2521 NPCDKKPYAPDKYIGR--ENPCENRENEFYDYEMKCYKNSKFOEKKRVOPRRER 2578
 Db 1361 GVCGLTLDAAE-FLNLKNGPCNKESEENKKADELDFKPPDTRKADN--CKP-- 1413
 QY 2579 MCLBNLDEIKIERLKDSNYLLKAVRTARNEGJDIIFNFSENGCAIMPICD--TKRYSFA 2637
 Db 1414 ---GSEFKIK-----CENHNCSSGGMTQCGDKCTTIANT 1445
 QY 2638 DLGDIYRGT---DMLRIGYLPPEYELKLVKVEFYIYIGKMNKRNKKNKNDVOTFBSAMW 2694
 Db 1446 EIEINKTKEVIMLVSDDSKATEK-----DGLSECKDKGIFGFI-- 1487
 QY 2695 DANRKDIWAKMTCKABEDAKLFRKGRMDGFERITL- IODKCGHKDPPVDYIYPOFRW 2753
 Db 1488 ---RKDEWEC-----GKVCVDICNLKKONIGESD--KKTITMK--ELL 1526
 QY 2754 TEMSEYCKALMELEKFKKSDCHCTSDRCKANDYDENKCEOCKTRCOEYKKNFVYLKWSL 2813
 Db 1527 KRWLETF---LEDYNKIKHKSHTCKNGK-----GSKC-----IKGCVKRW-- 1564
 QY 2814 FDIQSNKRYELBOPIYTYISTYDHOVNVOKLTPKSCSVSESEYELHESKLANKF 2873
 Db 1565 -----VOQKKEEMKQK-----ERNE-----QTKS 1585
 QY 2874 NENDSSNIRTAFAETEPKSYEACSTLPKSNPLDNCPTDQNKDGKELQTFPCSRND 2933
 Db 1586 KTSDEYFNKSFLETWIPK-----IAVYND 1610
 QY 2934 YDNLDNMNAAYVLNSSDQNGVLLIPRRRHLCSTRPITAYNRKGDKELKKKLLTSAFS 2993

Db 1611 QDN-----VILSKFGNSCG-----CSASALSTN-----GNEE-----1638
QY 2994 QGOLLOKTKSEELCFEAMKISYADYSDIITKDTMDMDSLSKIKKIETFSWEATEENRK 3053
Db 1639 -----DAIDCMIKKLEKKIDCKCKRPGEN--1662
QY 3054 TWMENNRQIWMHMLGCKYKATSKVTLDEGWCQLPKDEETNOFLRMLEMAKQACKKKH 3113
Db 1663 -----SGQTC-----1667
QY 3114 VSDSLTKCPRSNEDNEFEASELROPGCCONDIRKYSILNLIKNTMENLNIKYKQLKDS 3173
Db 1668 -----NELTLHP-----1674
QY 3174 SGNIDNKPEENVOYSIKSKDSQCALENDINEITVGTKNENNEKEVLKLYPGLY-F 3232
Db 1675 -----LDVODEDEPLEETEENP-----VGKHPSCECP 1702
QY 3233 VEDETHKNHVLDOGNKEE-QTVRPKALYFPHVDSFYQAPLFSHRAVQYDPKND--I 3289
Db 1703 VEDKK-----KEEGEICTPAS-----PAPAPAPASPSTPAPADEPEPTI 1746
QY 3290 LKSSIVYIYALGLIALHEMK-----KKFSSVDLRLILNIPG 3329
Db 1747 LOTTIFLGIALAGIAFLFKYIYICVYMYTMCPCITMKTKKHPVDFEYINIPKS 1806
QY 3330 EYCMPTLESKNRYIPYRSGPYKGYIYMEGDT--SGDEKYMMDLSSSDIT--SSESEY 3385
Db 1807 DYDIPKLSFNRYIPYTSQYRGRKRYIYLEDGSDGTSGYDHY-----SDITSSSESEY 1860
QY 3386 EELDINDIYVGSPPKTYLLEVLEPFSKRDPISDDTPS--NTPRTNRIIDEMWNLKH 3442
Db 1861 EEMDINDIYVGSPPKTYLLEVLEPFSGNNTTASDIQNDIQNGIDPSNKSNDENWNLKD 1920
QY 3443 DFVSQYLPNTEPN--NNYSADIPNTEPNITLYSDNPEKPTIISHDRDLYGKEISYN 3500
Db 1921 DFISNMLQNPQVNDYKSGDIPFMTQPTLYFDEKPEKPTITSIHDRMLNGEESYN 1980
QY 3501 INNSTNTNDIPMANRDSRGIDLINDSL 3530
Db 1981 VNNSTNSMD--PKYVSNNYSGIDLINDSL 2009

RESULT 11
T28634
variant-specific surface protein 7 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
Accession: T28634
Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Gulnet, F.; Herrfeldt, J.A.; Peterson, D.S.
Cell 82, 89-100, 1995
A:Title: The large diverse gene family var encodes proteins involved in cytoadherence an
A:Reference number: Z20487; M0ID:95330813; PMID:7606788
A:Accession: T28634
A:Status: preliminary; translated from GB/EMBL/DDDJ
A:Molecule type: DNA
A:Residues: 1-2182 <SUX>
A:Cross-References: EMBL:L42636; NID:9886379; PID:9886380; PIDN:AA75399.1
C:Genetics:
A:Note: var-7

Query Match 9 0%; Score 1743; DB 2; Length 2182;
Best Local Similarity 20.3%; Pred. No. 2.2e-73;
Matches 757; Conservative 339; Mismatches 678; Indels 1960; Gaps 126;

QY 13 GNAASLEGDAKSPITKESHKS--ARNVLER-----YAKNIRHPSKYAKEHVDLSKGLDK 66
Db 10 GSGGSS--SGGKGDTSRITIVYVSDAKDLDDVGEKVEYKAKND--AKKIIELAKGLMNT 66
QY 67 AEFRRGGSTPVNKHNYYPYPCNLDHKEHTNLRDYDN--LRHPC-----GRBONRFE 119
Db 67 ANGRSSTASSIE-----TCTL-VKEY-----YERVNGDGKRHPCKDAKNEDVNRFS 114

QY 120 DESEECG--NKIRNYKRNDAI--ACAPRRRHMDCKNLEALNDINTONIHDLGNVLTAK 177
Db 115 TLGGQCTYNNRIKDSQGDNDNVGACAPRRLHLDYNNESID--TTSTTHMLLEVCMAK 172
QY 178 YEGESIYNNH--PHKGT-----SDACTALARSFADIDYRGIDME-----KPNVDKY 224
Db 173 YEGNSI--NTHYTHQHQRTNEDSASQLCTVARSFADIDYRGADLYGYNKEKEQOKKL 231
QY 225 EFGLEVEYKKIHGMEDEVKND-----YNDPG--SGNYTKLRAMWNNYRNRKWEAITCDA 278
Db 232 EOKLKIDIFKKIH--KDYMTNGAQERYIDAKGDFEQLREDMTWNSRREYWKALICHA 288
QY 279 SYSGIYMOSESNTPLFSNKCCHGKQKVPNTNDYVQYLRWDEGECECRKNITLAK 338
Db 289 PKENAFYIKTACVNGKTCQCHCIGDVPYTYDQYVQYLRWDEGECECRKNITLAK 348
QY 339 VKDSRCDRLKLVCSHGHGCTTTMKKGLIHDNKTCDSTCKCYEVMGNOQAEFK 398
Db 349 LOKOCRDYEQNLKCSGNGYCTKTYIKKGLVIGEHCTNCSVWCMTETWIDNOKKEFLK 408
QY 399 OKERYEKET-----OSYLSNDNRKFVNNINSEYKQYERKLETOYATN 441
Db 409 OKRYETEISGGSGSKSPKRTKRARASSSSDDN-----GYSEFKYKKLEKVEYQDV 460
QY 442 DTPLNLTNEKRYCKG---GLPGEKIDITFNASADKGIPTFRSEYQVCPGCKGDCIKY 497
Db 461 DKFLKLTNKGICOKOPQVONERKADNVDFTNEKYVK--TFSTELCEPCPCMGLEGGPWP 519
QY 498 THRSNDNRHVNEDYKPRPGVKTNTIVLYSGNEGODITOKLEPNFNSPTNYKDNKNOK 557
Db 520 KYVGDKTCGSAKRTYPERK---NTIDLPVLVLPKSOQNLKTKYKFNFEKCAP--GGGQIK 575
QY 558 WEQYKND---ENINCKLEQNTIINDNPKIISFNFEWLYTYLLRDTIKMNDLKT 612
Db 576 WOCYDYDEHRSKNNKNCVGTWDFKQKQYKSYNVFEDWVHDLHDSVEKTELK 635
QY 613 CINNTTF---HCIDECNRNCLGDBRVKOKKEEEMNSIKKLTFTKKNIOQSYSNINLP 668
Db 636 CINNTNGNTCRNNKNTKCTDGCFCQKVEKQKQWMAIKHFGKQTDIVQO---KGLI 690
QY 669 EGYEFKVMKLDKDEAKWEKMEINIKRKNNEFSNLNNRYLE-----NAI 714
Db 691 VFSYGYLDLV---LKGNLQNIKQVHDGTDIDIRIKKILDEDAVAVYLGKDWTTI 746
QY 715 ELLDLKETATYCKDNNTNEACETSHNATNTPCVKRGSGTOPTKIKEIAQYKRSAYE 774
Db 747 DKLLQHEKEQAEQCKR--QEECE-----KKAQ 773
QY 775 EARRGLHKLKGAHEGITYRGRGRKDFKDNLCIMIKHSNRNLGFSNGPCDGGTGDGI 834
Db 774 ESRRRSAAE-----TREDER-----TOOPADSAGE---797
QY 835 QTRFVGTWEVDEPHRKDHEDVIMPPRRRHICTSLNHLQTDHPLNGIYVDLNNNS 894
Db 798 -----VEEEDDDDDVDEDED-----DDVQOE 820
QY 895 FLGDVLLSAAYEANKIIRMTKEKNMLGPREVTDPKHQTITCAIRYSFADIGDIINGRD 954
Db 821 EEG-----KEEGTVEYEVTEVEET-----842
QY 955 LWRNGDMVVLQGLHETVFGNIHKSLSKGNDRKNDAPKYLKREMMWENRKAQVEM 1014
Db 843 -----842
QY 1015 KCDIKYLLKDSHGSTQSSYGYSDHTPLDDYIPQKLMMTEMAEWYCKVQKREYDLKE 1074
Db 843 -----VTE-----OE 847
QY 1075 KCECKDKDNGQCGCTKESGTCCTCATECNEYNDIIGLMEQWNIISDKYKELHQAOMS 1134
Db 848 GVKRPC-----DIVKEL-----858
QY 1135 VNSGICFASSTAKNHNIDRNVIEFLSELVQONGSKNSGTSDESAVIGTMTYENVAYL 1194

```

Db 859 ----- 858
QY 1195 HDYGNFDDCOSONEFCDKSDGKDNKKAFAFRDKPODHDGAGCGCKSGSPRVRVQJTKTKKA 1234
Db 859 ----- 875
QY 1235 EFKTECKTVNDILKENDGKQVDBCHKRNKNGYPRMOCGNILV-----EDPRVCM 1308
Db 876 ----- 906
QY 1309 PROKLCVHFLA-----NDNEIKLOSov-----NKEKFIKSAAEFFSvWYY 1353
Db 907 PRRRRLVYGGLSOWASNGGDETTESVSEASAPSQSEKRLTAFAIFISAAIEFFELMHKY 966
QY 1354 K-----SKDG-----EGNELDKELKE-GKIPPAFLRSMTYFEGYRDLFLG 1393
Db 967 KEERKPPATODGAGLGVSLPEPPSPGEDPQTOLQOTGVIPDPLRQMFYTLADYKIDL- 1024
QY 1394 TDISKGHGBSKLEQJDSLFKNGDOKSPNGKTRQEWTEHSEHIEAMLCALVKGAKK 1453
Db 1025 ----- 1024
QY 1454 DDEFENYNNVFKSDKSTLEEFKRPQFLRWLTEMVYDYCTTRQKYLKDVOEKCKSND 1513
Db 1025 ----- 1024
QY 1514 QLKCDTECNKCCEDYVKKMKKKKEMIPODKYKDERDKKFRQHIGVMVTDYGTNAD 1573
Db 1025 ----- 1032
QY 1574 YLNRKFTASCGDKPGSASVYQNRNIOLEQAYYDADKHCCKTFIENDKYNISKDKC 1633
Db 1033 ----- 1047
QY 1634 KGLVKEAMTGAJWKONKGNPNYNNLKELTEDVLFPSRLRIGFHALDGYTDPVYDENG 1693
Db 1048 KNLVLEAS----- 1055
QY 1694 LRRRLMEVATEGYNLGOYYKKEKKEKIKTSDAHKYSIEVPPCSAMKSYFDLBDITLG 1753
Db 1056 ----- 1066
QY 1754 IDNLEDEKQTEENLKIFPNKNGTSYVKGSDSTGNGPSTARKFPNNENKBEVNNAMICG 1813
Db 1067 ----- 1113
QY 1814 --YKRGRODNGSNGSARSDDELKKGVSPPSDDVPMGNRDEGTAYOFLRWFPAEWGEDFC 1871
Db 1114 LTYKE-----NDAR-----GTS----- 1125
QY 1872 KHEKELEKLVGACNDYCGDNEDARKKCTDAGYKFFISEMKQYKQIKYGENKDK 1931
Db 1126 ----- 1133
QY 1932 IYSEHPVAKDAEDAREYLDKOLKIKCENKSGDCEYCKAKDVSTORLTDGSONNPASLDD 1991
Db 1134 ----- 1147
QY 1992 EPKEVEGKCNQVPRGPRVREPRSPVSLISKATASKAKTAAPTQKPKVENLITE 2051
Db 1148 ----- 1147
QY 2052 MRAOTRRRAAQTRKRTSTATTESDVGTWVKAILSNKPSRSGISGCKNKTGYQPKW 2111
Db 1148 ----- 1147
QY 2112 GCIVGSKENENGICMPRRKKLGINNIQVINYTEENKRNNDIKAFIKCAALFTOPLML 2171
Db 1148 ----- 1159
QY 2172 KYIENPAENELONGTIPDEFKRIMTYTGYDYDMFEGTDISNDKILITVNSVTIILN 2231

```

```

Db 1160 ----- D 1160
QY 2232 ENNRKQDKKDELEKIFEMKNNKFIWEGMIYGLYHILTDENEKEKINDYNDMYKL 2291
Db 1161 ESGAKSND----- 1171
QY 2292 TBSLEFVYKRPQFLRMFTMAEFECNRRKREOLKLEAGC-----KEYCNGSNDK--- 2342
Db 1172 PPLTKNFVEIPTEFFRWLHBMGNSFCFERAKRLAQJLHCEMDDEGKQY-----SGGEVCE 1227
QY 2343 -----TOCAEACVYONFETIKMKTELEERQREKREKRDGKQKYYIDYSTERD 2389
Db 1228 ELPFSKQYVNILODLSSCAPCRLIYTWLEKKTEKEKQKAYEQDQ-----SNYENKOD 1282
QY 2390 IEKATCAHEYLLNMLKLELGNDCSOMQPSQLEPRTYQOSQSDANDMPESIDVP--E 2447
Db 1283 -----KC-----QTSNNANANFSRTLGASPIAA 1306
QY 2448 EFNKQCPBELSKSGSMIHKKITTEPKIPANCYERAKAYILSKAEANNMDITLKEFPIES 2507
Db 1307 EF-----LOKUGS-----CKNDGY-----ENGEDNKIDFK-NPDKT 1337
QY 2508 TREKESKSWTNNNPCDPKPYAPDKYIGRRNCPENRENFKYDYEMKCKYKNSKFTQEK 2567
Db 1338 FKEAHS-----CDP-CRITGVK-----CONGH----- 1358
QY 2568 KRVCPYPRREHMCRLNLDLKEIKERLSDSNYLLKAVRRTAREGIDILKFNSENGCAMP 2627
Db 1359 --CVGSANGRECKNN--KITAEIDKN-----KIDPAGNITEMVSDSTN----- 1398
QY 2628 ICDJMKYFADIGDVRGTDMLRIGGYLPVYEIKLYAFEYIYKMKNNKGNKKNKYDVQ 2687
Db 1399 -----TFEHLGD-----CGSSGIFGI----- 1415
QY 2688 TFRSAMMDANKRDIWK-AMTKAPADAKLFRKGMDFERTILLID-KCGHKDDPVDY 2745
Db 1416 -----RKDEMKCANVC-----GVDICTLEKKIKNOGED--KKY 1447
QY 2746 IPORFRMTESEYYCYALMELEKFKKSCDHC-KTSDRCKNDYDEKCKQCTRCQERY 2804
Db 1448 ITMK-ELLKRLLEYF-----LEDYMRIRKKIKLCTKCKDGCK-----CJRGJIE-- 1490
QY 2805 NEVLKMSLFDIOGNKYKLEYEOPITYKISTYDVQNFVOKLTFKSECSVESFSEILHE 2864
Db 1491 --KW-----VQ-----EKTEMO----- 1501
QY 2865 TSKLANKFNENOGSSNIRTYAPETPKSYKZACSTLPSKNPLDNCPTDNRKGEKQ 2924
Db 1502 -----KIND-----TY-----LE 1509
QY 2925 TFFPESKNDYDNMLDNNAVYLVNSSDNDKGVLLIPRRRHLCRPIITAYNRYKGDKEIJK 2984
Db 1510 QY-----KNDGNTLTN-----LEOFQYR----- 1529
QY 2985 KLLTSAFSOGOLGQKYKSEEBELCEAMKYSYADYDIKGTDMQDLSKIKKIFET 3044
Db 1530 -----TEKNMINKPDGID----- 1543
QY 3045 SNEATENRKTWMENNRQIWMHMLCGYKATATSVTYLTDEGMCQLPKDEETNOFLRMLIEMA 3104
Db 1544 ----- 1545
QY 3105 KOACKEKHVSDLSKTKPCRSNEDNEFASSELLROPCONDIRKYISLNTILKNTMENLNI 3164
Db 1546 -----KTSQGLNSTDN-----SQNSNNND-----LVCLCLNLOK 1575
QY 3165 KYKOLDOSSGNIDNKPSEBNOVSYIKSDSOCALBLNDINELVGTKNENN----- 3217
Db 1576 KISECKEJQSGOTQ--PCQ--NSSLSGKESTLVEDVDYDYE-----ONPEKRYEQPKFC 1626
QY 3218 -EKEVLYKLYLGLYFVEDETHKNHLDGNI--KEEEOQVRKALFYLPPIHVDYSYQAPL 3274
Db 1627 PDWKPEPKENDEVGTGCGDEKKEKRYVEDSVIQRKEEBAASAPESPLTP-----EAPK 1680

```

```

QY 3275 FSHRVAQOYD-----KN-----DLKSSIVYVSAGL-----IALHEMKKKKS 3316
Dy 1681 KEENVAVERPPPPPKRRKIKRIRNLDHPAIVIPALMSTJMSIGIFPAFYFYLLKKTKS 1740
QY 3317 SV-DLRLINIPGEGVMPLETSKNRYIPYRSGPVKGYIYMEGDTSGDEDKYMMDDLS 3375
Db 1741 SVGNLFQLOLIPSDYDIPFLTKSNRYIPASORHGKYYIYMEDDSSGDE-KIAFMSDT 1799
QY 3376 SDITSESEYEELDINDIYVPGSPKKTLIEVLEPSKRDIPD-----DTPSNRPR 3428
Db 1800 TDISSSEYEELDINDIYVPGSPKKTLIEVLEPSKRDITONDINDIPSDIPMSDTP 1859
QY 3429 TNREIDDEMWELKHDFVSQYL---PTEBPNNYKSADIPMTEPNTLYSDNPEEKPIIS 3485
Db 1860 P--ITDEDMQKKDFISNMLQTOQTPEIRLHDVND--NTHHP--TMSHNNADOKPFI MS 1914
QY 3486 IHDRLDTYTGKEIST-----NINMSTN-----TNNDIPNARNDSTRGIDLINDS 3529
Db 1915 IHDNRLPESGEYVYNDMENSNNPINDISTDSTNSMDLSTNNHSPYDKNDLXSGIDLINDA 1974
QY 3530 LVVNLILLY--MOK 3541
Db 1975 LSGNHHIDYDEMLK 1988

```

RESULT 12
T14029

variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
N:Alternate names: erythrocyte membrane binding protein 1 (EMBP1)
C:Species: Plasmodium falciparum
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
C:Accession: T14029
R:Chen, O.; Barragan, A.; Fernandez, V.; Sundstrom, A.; Schlichterle, M.; Sahlen, A.; C
J. Exp. Med. 187, 15-23, 1998
A:Title: Identification of Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1)
A:Reference number: Z17860; MUID:98080592; PMID:9419207
A:Accession: T14029
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-2228 <CHE>
A:Cross-references: EMBL:AF003473; NID:g2961467; PID:g2961468; PIDN:AC05730.1
C:Genetics:
A:Note: FCR3SL.2-var1

Query Match	8.7%	Score 1689.5	DB 2	Length 2228
Best Local Similarity	20.1%	Pred. No. 7.2e-71		
Matches 756	Conservative 355	Mismatches 661	Indels 2015	Gaps 137

```

01 33 KSRNVRIRYAKNRI RHPSKY--AREHNDLSKGLDTKAFEPFGGPSTPYPNKNANYYPYCNL 90
02 13 EDKAVHVDLEFQOKY-HDEVEHAEANVYSEKJSLSLSTIGETAFYK----- 59
03 91 DHEKHTNLRVDV---NLRRPC---HGREONPFDEDESECGNKIRNYKRR-NDAIAC 141
04 60 --SMOTSEKYTELLEANSKRPPCKKDGKNDVDFSVKEQAGIDYK---KMKCSNMTC 113
05 142 APPRRRHCDKNLEALNDINTONT-HDLGNVLTAYTEGESIYVNNRP-----HKTS-D 194
06 114 APFRRLHJCNNFPMNMSNDSSKAKHDLAEVCAAAAYEGSITHTHPKDKSKYPSGDF 173
07 195 ACTALARSFADIGDIVGIDKF-----KRVNHDVYETGLREVFKKIHGDMED-EV 243
08 174 MCTMARSFADIGDIRRGRDYLGNKKKKONGKETEERKELOKLETKFIHDLKJKEA 233
09 244 KNDVAPDSSGVYKLRKMMNVNKNKYWEALTCASY-KSIFYMQSSNT---PLSPNPK 299
10 234 QKRYGDDDPNFYKLRBDMWTANRETYGAMTCSEKELDNSSYFPATONDNGOQSPQTHNK 293
11 300 C-----GHKOK-----VPTNDVYPOYLRPFDEMGEECEKRRNITKTKVYDSC 343
12 294 CXCDCDKCANNGKPPACGDVYTVPTFYDYPOLYLRPFEEBAEDPCRRKKKTKENLEKOC 353

```

OY	344	R-NK-----	ERLYCSNNGDCTTTIMKGLIHLNDKCDJCSGTCKKVEFWYVLGNQOEAFFKOK	400
Db	354	RKCKSDERCSNNGYDCBQTISRKGVMGKGCDFCEPACHSEYMWLDNRKOFDOK	413	
OY	401	EKEKEJLOSYSNDNKKFVNIN--	SEYKOFYEKLETOYATNDFNLINBGRKCK---	455
Db	414	-KYKEISDGGRRKRAVGSTTKYEGEKSFYKLLKNDGYVDAPLGLINNEKCKRIT	472	
OY	456	-----	GGIPKEDITFTNSADDKGIFYRSEYCOVCPDCGKCDGIRY	497
Db	473	DGKINFEVNSGGGVGGSGGSGASGTND-ENKGTIFYRSBQCOCPDCGVQ-----	525	
OY	498	THKSDNDRERANNEDYPPMG--	KPLPN--ITVLYSGEODG--ITOKLENFC--NS	546
Db	526	-HKGNOMER-KTYVKTKMRSKLYKPINGKMYLLLSLVVYDMMLKKNMKEPCLTONS	583	
OY	547	S-----	TNYKDKNOK-----WECYKDEINRCKLSDONTINNDD-----	581
Db	564	SDGVSQVYTTGASGSGSEKKELYDERKC-YKINERYOKNVOGEVEBDEDLKAGGICLI	642	
OY	582	-NPK-----	IISEFPELWYVLYLBDITKMDK-LKTCINN-TTTH	620
Db	643	LPNPKKNKEYSEAKSQNNHADIOKTFHDFEYVYVAMHLDOSIHMRTKRLKSCISDGTKMK	702	
OY	631	CIDECNRKICLFDWVWVQOKEBENSISKLPFTKKNKIQOYSYNSINNLFGCIFYKMDKLD	680	
Db	703	CRGCKNKKCDCEFEKVMQKETEKPIDKHFKTOEGIEPEGY-----	743	
OY	681	KDEAKWEIMENIKRKNESNLENNNDYLENAIELLDHLEKETAICKDNNTNACETS	740	
Db	744	-----	FTITLE-----LILKLOFLBE-----DTEBNTENS	767
OY	741	HNAATTNCPVKPRGTOPTKNIKEIAOYFKRSAYEAERNGHLKLGKAHEGIYKKGRRK	800	
Db	768	IDA-----	IIIEA-----	773
OY	801	DFKDNLCRIMIKHSNNRLGFSNCPBCKGKGTDGIGIOTRFVVGTEWEPDPEHMKRKHEDYIM	860	
Db	774	-----	-----	773
OY	861	PPRRRHICTSLNEHLORTDHPRLNGINIVDDLNNNSPLGVDLISAKYEANKIIMYEKENNL	920	
Db	774	-----	BEKLHQ-----KILLENE-NNL	791
OY	921	KGPKEVTDPKHOTITCRAIRYSFADIGDIIRGRDLWERNGDWYKLOGHLETFVGNHRS	980	
Db	792	AVVNAAGTE-QKTL-----	MDKLLNH-----	810
OY	981	KGKGNDKYNDAPKYLKLRNMMWEANRAKYWEAMKCDIKYL-KDKSGHSTOSSYSGYS	1038	
Db	811	-----	ELN-----DATKCKDPLPEEDKRSGRS-----A	834
OY	1039	DHTEPLDYIIPQKLMWTEMAEMWCKCYQKKKEYDLKLEKCECKDKDNKGCGCTESGTGCTK	1098	
Db	835	DPSR-DIFTP-----	RPEKCHDE-----	852
OY	1099	CTEACNEYNIDITIGLKEQMMNIIISDKYELHEQOQMSVNSGLEASTAKNHIDRNVIEFL	1158	
Db	853	-----	NEDD-----EDVARDDEBETAKETTESADDTTSL-----	883
OY	1159	SELYOONGSKNSKSTDESASVIGTNTTYENGVAYLHDGPNFDDCOSONEFCDEKSDKGD	1218	
Db	884	-----	DVCPPIVGKVLV-----KD	896
OY	1219	NEKYAFRDKPODDHAGAGCSGSGSPFPVJOIKTKKKAKEKDTECKYVNDLILKENDKQOYE	1278	
Db	897	NESL-----	OD--ACSLKYGGNSHLAGRCVTPSGEPTT-----	928
OY	1279	DCHEPKNSNGYPDWQCNINILVEDPBYCMPPRROKICVHFL-----	ANDNEIKK	1327
Db	929	-----	SSDKNG-----AICYPFRRRRIKLYIKYDMATKTSPVQASGSEASS	969
OY	1328	L-----	QSOVNLKEAFIKSAAAEFTFSWYTK-----SKDGEKNEL-----	1363

Db 970 TSGSTPPDSEKELKAFVESAAIEFFLWHRYKEEKAAVAGAGHGLPVEGSESPED 1029
Qy 1364 -DEKLEKGIIPAPLSEMEFTPDYDRLFTGISKHGSGSLKEIDSLFKNGDQKSP 1422
Db 1030 PEKIKRGKIPDGLKRMFTTLDYRDILF----- 1059
Qy 1423 NGKTRQEMWTEHSHEIWEAMICALVKGAKKDDFTENYGNVYFSDKSTTLEFAKRPQ 1482
Db 1060 ----- 1059
Qy 1483 FLRWLTWYDYCYTRKYLKDVQEKCSNDOLKCDTECNKCEDYVYTKKKKKEMIPD 1542
Db 1060 ----- 1059
Qy 1543 KYKDERDKRRFDROHIGVWYTDYGTNADYLNKRFKASCGDPGASVYQNRITQLEK 1602
Db 1060 ----- 1077
Qy 1603 QAVYDADKHCGCKFTIENDDKYTNISSKQCKGLVEANTGAIKQWONKGPNNYNLKELT 1662
Db 1078 ----- 1084
Qy 1663 EDVLPFSRLRICFHALDGNYTDEYKDENGLRKRLMEVAATEGYNLQYKEKEKEKI 1722
Db 1085 ----- 1098
Qy 1723 KTSDAHKIYEVPCSAKMSFYDLIDILGIDNLEDEKQTEMLKKTFFNKGTSGVK 1782
Db 1099 ----- 1115
Qy 1783 SDSTGNPSTARKFPFNWNEKCEVNNAMICYKRGDRDGNCSNARSDEDLKCKGSPSD 1842
Db 1116 APNLVSHP-----QIWEENNGKIWHMGVCL----- 1142
Qy 1843 DDYPMGNDEGTAYOFLMFAEMGEDFCRHEKELEKLVGACNDYTCGDNEDKKCTD 1902
Db 1143 ----- 1142
Qy 1903 ACTQYKFISEKPOYEKOIKYGENKDIYSEHVAKDAEDAREYLDKOLKICENKSG 1962
Db 1143 ----- 1153
Qy 1963 DCEYKCMKVSTORLTDGNSONMPASLDPEKVEKGCNCOVPRGPVHRETPSPRSL 2022
Db 1154 ----- 1153
Qy 2023 ISKATASKEAKTAPPTQOPKKEVNTLTEMRAQOTRRRAOQTRKRISTATTTESDVTM 2082
Db 1154 ----- 1161
Qy 2083 VKALISNRPDSRGIEGCPNFTYGOYPKWGCIVGSKENENGICMPRRKLCINNIOYL 2142
Db 1162 ----- 1181
Qy 2143 NYEENKRDNDIKAEFIKCAIEFOFLMKYIENPAENELQNGTTPDEKRIYTYG 2202
Db 1182 NVKLD----- 1197
Qy 2203 DYKMFEGTDSNDKLIITVNSYTTINENKCKQDKKDELRKIFWENKKEFIWEGM 2262
Db 1198 ----- 1208
Qy 2263 IYGLTJHLTDENEKEIRDNQYNDMTKLPSLEEFYKRPQFLWFTENAEEPCNKKEQ 2322
Db 1209 ----- 1336
Qy 2323 LKLEACCKEYECNG-----SNDGKTOE----- 2360
Db 1237 LKQITKVDCKYE--NGDVGRSGDGEACDSISTHDYSTVPFNCBGGCKHCSSTRKWIERR 1294
Qy 2361 KTEHEROREKFKKXKDKKXKDPSTERDLEKATCAHEYINMKLKEJCGNKDCSCWQKPS 2420

Db 1295 KIEFKQSNAYQ-----OKTATRNGNTFDEKFC----- 1325
Qy 2421 SOLPRTQOSSDANDMPESIDYVPEEFNKCECELSKSGSMHTKKTIBPKIPMNCVE 2480
Db 1326 ----- 1333
Qy 2481 KAAVYLSKAEANNMITYLKEKFIPIESTKEKSNWNTNNPCDPKPYADKYIGRRNP 2540
Db 1334 -AKFLER----- 1355
Qy 2541 CENRENEFKVDYEWKCYKNSKFEYKRCVCP--PREHMLRLNDEKIERLDSNVL 2599
Db 1356 ----- 1382
Qy 2600 KAVRTARNEGIDITIKNSENSENG-----CAMPICD----- 2644
Db 1383 ----- 1420
Qy 2645 GTDMLRIGGLPYVEIKLYKFEYIYGKWRNKNKGRNKNYNDVOTF-----RSAMWDAN-- 2697
Db 1421 VSD----- 1444
Qy 2698 ---KRDYKAMTCAKABEDAKLERGRMDGEF--RITLIDCKGHKDDPYDDYIPORF-- 2750
Db 1445 KGRDQVWK-----GYCGVDICGQTNINERTDK-----EYIQIALF 1484
Qy 2751 -RWMTWSEYCYKALMELEKFKSCDHC--KTSDRKNDYDENKCEOCKTRCOEYKNF 2806
Db 1485 KRWENF-----LEDYNNINDKISHCIKKGSGCINGCEN--SKLEWIEK-- 1532
Qy 2807 VLKRSLEFDIOSNRYKELYE--OPIYKISTYDHOVNOFKLTFSECSVSFSEYLYH 2863
Db 1533 IAEWENI-----KRFNDQYENKODPDYVNSI--LEELPKIYVVDQNVKILCYE-- 1583
Qy 2864 ETSCLATKFNENDGSSNIRIYAFEEPTKSIKACSGTLPKSN--PLDNC--PLDQNKDG 2919
Db 1584 ENSKGTILISNTONKN-----DAIDMLKKLYKAKANCGKPSGEMQSD 1629
Qy 2920 CKELOTFTCSKNDYDNLNLMNAYLVLNSSDQNGVLIIPRRRHLCSTRITLYANTRKD 2979
Db 1630 CKEPPL-----PDEEDN-----PEENTLEPK--PCP--PTTQPPBEKG-- 1666
Qy 2980 KEILKRLLSASOGOLCOKYKSEBELCEFAKMYSYADYSITIKGTMDMPTSLSEK 3039
Db 1667 ----- 1680
Qy 3040 KIFETSNEATENRKTWENNRQIWMHMLCGYKIATSKYTLDSGMCOLPDETNQFLM 3099
Db 1681 K--BESSE----- 1693
Qy 3100 LIEMAKOACEKKHVSDSLTKCPRSNEDNFEASELLRQPGCONDIRKXYSILNLIKMTM 3159
Db 1694 -----PAEEPPAPTAESEET-----ETNFP--EPPG----- 1717
Qy 3160 ENLNIYKQJAKDSSGNIDNKPSSENVOSYISKDSOCALBLNDINEIYTGKNNENEF 3219
Db 1718 ----- 1728
Qy 3220 KEVLKRLPYGLYVEDETHKNHVLGNKEEBQYVRKALYFTTPHYDSFYAPLESTHR 3279
Db 1729 PPT----- 1749
Qy 3280 VAOYDPRKNDILSSISVIVTALGLTALHFMKKKFSV--DLRLILNIPQEGYGPPTLES 3338
Db 1750 ----- 1800
Qy 3339 KNRYPYBSGYPYKGYIYMEGDTSGDEDKYMDJSSDITSSSEBYEELDINDIYVGS 3398
Db 1801 SNRTIYVDRKKGKYYIYMEGDS--DEDKYAFMDTDYVTSSESEYELDINDIYVGS 1858
Qy 3399 PKYTLIEVLEPSPKRDIPSD--DTPS-----NDTPRTNFIDENNELKHDVVSQYL 3449
Db 1859 PKYKTLIEVLEPSPGNNTTASGKNTPSDTRNDIQDNGIPSSKITDNEWNLKKEFTSNML 1918

3450 PNTERN---NNTKSDIPNTE-----PNTLYSDNPEKPEITIHIDRLYT 3493
1919 QN-ONDVNDYSGNSTNTTTSRHNVDNNTTMSRDNMEENLLPSIHRLNYS 1977
3494 GKEISYINMSTNNIPMARNDSYRGIDILINDSL 3530
1978 GEESTYNNM-VNSMDPIFNDDNNVSGIDILINDSL 2013

RESULT 13
C71625
variant-specific surface protein 1 homolog PFB0010w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: C71625
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Partee, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
Accession: C71625
Reference number: A71600; MUID: 99021743; PMID: 9804551
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1711 <GAR>
A:Cross-references: GB:AE001366; GB:AE001362; NID:93845070; PIDN:AAC71792.1; PID:9384507
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0010w

Query Match 8.1%; Score 1574; DB 2; Length 1711;
Best Local Similarity 17.6%; Pred. No. 1.3e-65;
Matches 635; Conservative 259; Mismatches 496; Indels 2218; Gaps 83;

33 KSARVLEKRYAKNI-RHPSKYAKENHVSIDLTAKFEFGSGSTPYNKHNYYP----- 85
13 ESKVNDLPDIDGKVKYEKTEKIKRYTELHGDLSKATY-----PNDKH-----PEGSTEN 62
86 YVCNIDHKHTNLRYDDVNLRHPCGREGQNRDEDESESGNKRINRYKKNALACAPPR 145
63 NCKKLOYDYNINVTN-GFGQEPCEPDYIERSDTEGACQDKR---KIDNSEGACAPYR 118
146 RHHMDKNLEALNDIN-QNIDHDLGNVLTAKYGESEIVNNHP-HKGT-----SDACIA 198
119 RHVVCVRNLENINDYSKINNHNLLVEVCLAKYGESEITGRYPOHQTNPDTKSQDLCTV 178
199 LARSPADIGDIYRGIDMEK-----PNVADKYETGLREYFKIHDGM-----EDEVAN 245
179 LARSPADIGDIIRGNDLYRGNGTKKKKKKKKLEBNKTIIFGHIYDELAKNGKTNGEBELQK 238
246 DYNPDGSGYTYKLRAMANNVNNKWEALITDA-SYKSGYFMOSSENPDLFSPNPKGHRK 304
239 RRGCKNDNFYOLREDMDMANREYWKAITCNAGSYOYQPTCGRGEIPYVLSKQCIA 298
305 GAVPTINDVVPQYLRMEDEMGEEPCRRNIKIKKYKDSOR---NKKERYLCSHNGHCTT 361
299 GAVPTFDYVPOYLRMEDEMAEDFCKKKKKKIPNVTNORQYORGERK-YCDRDGNCDG 357
362 TIRKQYIYRLDTCRKSIACTPAEMIDNOKQEPDKOKOKONELSGGGGRKORSHS 417
358 TIRKQYIYRLDTCRKSIACTPAEMIDNOKQEPDKOKOKONELSGGGGRKORSHS 417
422 NSE---YKQYETKLEKTOYATND---TFLNLENGYCKGGDL-GEKDIPTFNSAD 472
418 TREYDEYEHFNEELR-----NEGKDVNSFLQLSKERIKERLOVGEETANYGFEENE 471
473 KGIYRSEKCYQCPGQGVKCDGKITKHSND-REVVNNEDEKRPMPGAVPTNITVYLSGN 531
472 SUTFSHTEICDRLGCVGSSDNCRRKPKDKSCDEQITTKETPPE---NTTKIPKLTAK 528
532 EOGDITQKLENTCNSYTKKRN---OKWECY---KDE---NINRC-KLEONTIN 579
529 RYGIILKYEKFCNKS---DGNNGQILKKWCHYEKKNDKDGONGDINNCGIDMWTSKN 584

580 NDNPKIISFHNFEELWVLYLLDITKMNDKLTCTCINNNTT-HCIDEONRNICLDFRMVQ 638
585 YVPP--ISYSEFYGSIIIDMLNESIEWERLKSICINDALGCRGCKNPKCEYKRWYK 642
639 KEENMSIKKLTTKKKNIOOYSYNNINLFEQYFFVYMDKLKDEAKKMELENKRRKN 698
643 KKEWDRIKEEPRKQDL---LKDIAGMAG-----ELLE----- 674
699 EFSNLENNRDYLENATEILLDLKETATICKDNNTNEACETSHNATNCPVPRGQPT 758
675 -----FYLEN---ILEDMKN----- 687
759 KIKETIAYFKRSAYEARNRGLHLKKAHEGYKRGRRDPKDNLCRIMIKHSNRL 818
688 ----- 687
819 GTSNCPDCKGQDGIQTFVVGTEWVPEPMRKDHEVIMPPRRRICTSNLEHLOT 878
688 --ANG-----DP----- 692
879 DHPNGNIVDVLVNSFLGDLVLSAKYEANKIIRMYKEKNLKGPREYTPKHQTICRA 938
693 -----KYLEKFE----- 700
939 IRYSPADIGDIIRGNDLWERNGDWYKLGHLETFEGNIHKSLSKGNDRYNDAPRYLKL 998
701 -----ILGKRENEVQPP----- 712
999 RENWMEANAKYWEAMKCDIKYLKDSGHQSTQSSICGSDTIPLDYIPQILRMKTEWA 1058
713 -----LTKRK-----TTDDFLEKEL----- 727
1059 EWCYKQKREYDKLEKCEKCKDKNDGCGTKEGSGTCTEACNEYDIIGIMKEQKN 1118
728 -----NEANCVKEKPNDCPRKQKAG----- 749
1119 IISDKYKELHEQAQNSVNSGIEASSTAKNHDRIYIEFLSELYOONGKSNKSGTSDS 1178
750 -----DGAAPSDPPREDITH-----GHEHSSD- 773
1179 AVIGNTTYENYGAVYLDHGNFDDQSONFDEKSDGKNKRYAFRDRPOHDGACGK 1238
774 -----DEEEEEEQPPAEGTEGEE---K 796
1239 SGSKPTRVQIKTKKAEKQTE-----CKTVNDILKENDGKROVEDCHPRKNS 1286
797 SESK-----EVNQOETPRQDKETKYPTTPYDVCDYVKTALADGSLMAA---CSIKYVT 850
1287 NGYPMQC---GNINLVEDPRVCMPPRRKOKLCVHFLANDNEIKILO--SOVNLKEAFIKS 1341
851 GRNYGWRCIAPSGTSGKGAICVPRQELCLYVL-----KELSDTYQKGLREAFIKT 904
1342 AAEEFESVYVYK-SKDG--GNELD-----KELKSGKITPAPLBSMFTPDYRDLFG 1393
905 AAQETVLLQKTKEDQNTASTELDIDPOTQLNGEITPEDFKRMFTTFEDYRDLFG 964
1394 TDISKGGHGSLLKEQIDLFKNGDQKSPNG---KTROEMTTEHEHEIMELCALVYI 1449
965 RYIG---NOLDVYNNNTITAVFQNGH-IPNGQKTDROGRBEPFGTYKDKIMKGLCALDRA 1020
1450 GAKKDDFTENYGNVYKFS--KSTLLEEFARPOFLRLMLTEWYDYCYTRKYLKDOE 1507
1021 GGRK-TLLETVYNSVTENGHLTGTLNFAFSPLRMTTEWGDQFCRERITQILOIKE 1079
1508 KCKNSDQLKCDIECNKCKADYVYKMKKKKEMIPDKYVYDEBDKRFDPQHGIVMTDT 1567
1080 KC----- 1087
1568 GTNATDYLRKFTASCGDRGSASVYQNRIOLEKQAYVADADKHCCTFIENDKYTNI 1627
1088 G----- 1088

QY 1628 SSKDKCGLYKANTGAIKMONGPNPNYNNLKELTEDVLPSSRLRICFHALDGNTPDE 1687
Db 1089 ----- 1088
QY 1688 VKDENGRLRLMEVAATEGYNLQYYKEKEKEKIKITSDAHKSYEVPCSAKMYFYDL 1747
Db 1089 ----- 1088
QY 1748 RDIILGIDLEDEROKTEENIKKIFNKNGTSVGKSDSTTGNDSTARKFENNENECVA 1807
Db 1089 ----- 1093
QY 1808 NAMICGYKRRDGDNGNSARSDEDLKKGVSVDYDPMGKNRDEGTAYOFLRWEAENG 1867
Db 1094 ----- 1093
QY 1868 EDFCKHEKELELVAGANDYTTCGDNEDKRRKCTDCTOYKRFISEMKPOYEKOIKKYE 1927
Db 1094 ----- 1126
QY 1928 NKDKIYSEHPVAKADADAREYLDKOLKJCEKNSGDCEYCKMKDVSTORLTDGNSQMPA 1987
Db 1127 ----- 1126
QY 1988 SJLDEFEVEGKCNCOVPRGPPVRRETPSPVSLSKATASKKEAKTAPTKORRYEN 2047
Db 1127 ----- 1126
QY 2048 LTTERRAOTRTIRAAOQTRKRTSTATTESDVTGTVKAILSNKPDPSRGJESGPNRTYGO 2107
Db 1127 ----- 1132
QY 2108 YPKWGIYKSKENENGICMPRRKKLCINNIQYLYNETENKRDIDKALFKCAIETO 2167
Db 1133 ----- 1142
QY 2168 FLMLKYYIENPAENELONGTIPDEFKRMYYTYGDKDFMGDTISMDKLIITVNSVT 2227
Db 1143 ----- 1147
QY 2228 TLINNNKKKODKKDELRKTFEKKKKFKTWECMITYGLYHLNDENEKIRIDNYOYND 2287
Db 1148 ----- 1153
QY 2288 MTKLPSTLEEFYKRPQFLRMFTENAEEFCNKRKBQLKLEAGCKYEGGNSDGTQCA 2347
Db 1154 ----- 1163
QY 2348 EACVYQNFYIKKWKTEYEROREKFKRDKGKRYKDYSTERDIEKATCAHEYLNMKLKEL 2407
Db 1164 ----- 1166
QY 2408 CGNKDCSCNOKPSSQLPKTQOSOSDANDMPESDIYVPEEFNK-CECPPELSKSGSMITH 2466
Db 1167 ----- 1212
QY 2467 KKTTEPKTPMNCVKAAYYSKEAENNDITLKERFPIESTKEKESKNSWTNNPCDPK 2526
Db 1213 ----- 1220
QY 2527 KPYAPDKYIGRRNPCENRENEKRVDEMKCYKSKFYQEKKKRYCVPRRRHMCRLNIDE 2586
Db 1221 ----- 1220
QY 2587 IKIBRLKDSNYLTKMVRRTAREGIDILIKFNSENCGAMPICDTMYKSPADLDIYRGF 2646
Db 1221 ----- 1220
QY 2647 DMLRIGLYLPPVEIKLYKVEEYIYCKWRNKNKGNKYNDVOTERSAWANDNRDINWAKMT 2706
Db 1221 ----- 1220
QY 2707 CKAPEDAKLFRKGRMDGFERITLLQDKCGHKDPPVDYTIPOFRWMTWSEYTCALME 2766

Db 1221 ----- 1220
QY 2767 ELEKFKSCDHCKTSDRCKNDYDENKCEQCKTRCOEYKNFVLKMSLPDIQSNKKEIYE 2826
Db 1221 ----- 1220
QY 2827 QPIYTKISTYDHVQNFVOKLTKFKSECSVESFSEYLETSCCLNFKFNENDGSSNIRTYA 2886
Db 1221 ----- 1220
QY 2887 FEETPKSYKACSTLPSEKPNLNDNCPDQNDGCKELQTFPCKSKNDYDNNLDMMNATLV 2946
Db 1221 ----- 1223
QY 2947 LMSDDNKGVLIPRRRLCTRPITAYNRYKRGKREILKLLKLTSAFSOGQLLAGQKYSSE 3006
Db 1224 ----- 1226
QY 3007 ELCFEAMKYSYADYSDILKGTDMMDTSLSEKIKKIFETSNEATENKRTWENNRRQIWA 3066
Db 1227 ----- 1226
QY 3067 MLCGYKATISRYTLDEBGCQLPKDEETNOFLRWLLEWAKQCKEKKHVSLSLTKCPRSN 3126
Db 1227 ----- 1229
QY 3127 EDNFEASELROPGCQNDIRKYISLNLIKNTMENLAIKKYKOLKDOSSGAINDKPSSENV 3186
Db 1230 ----- 1232
QY 3187 QSYTKSDSQCALBLNDINEIYGTGKNNNENEFKVLKLYPGLYFEDETHKNHYLDGN 3246
Db 1233 ----- 1236
QY 3247 IKEEOYVRPRALYFTFPYHDSFYQAPLFTSHRVAQDPKNDILKSSISVYISALGLIA 3306
Db 1237 ----- 1263
QY 3307 LHFMKKFKFSVDLRLINIPQEGYCMPTLESKNRYIPYSGYKKGKTYIYMGDT---S 3363
Db 1264 YFYIAKKTITIDLKRIYINIPKSDYDIPYKLSPPRYIPYSGYKRGRIYIYLGDSGTDS 1323
QY 3364 GDEDEKYMDLSSSDIT-SSESEYEELDINDIYVPGSHKYTLLEVLLEPSKRDIPS--DD 3420
Db 1324 GYTDHY-----SDIYSSSEYEELDINDIYVAPBAPKYTLLEVLLEPSGNTTASGNN 1377
QY 3421 TPS-----NDPRTNRFTIDDEMNELEKDFVSQYLPNTEPN---NNYSADIPTNTEPN 3470
Db 1378 TPSDTONDIONDGPSSKITDNEMNWLKDEFISQYLOSDQPNVDPVYSSGDIPLNTQPN 1437
QY 3471 TLVSDNPEEKPTIISHDRDLYGKEISYINNSTVNDNDIPNANNDYRGIDCLINDSL 3530
Db 1438 TLVFDNPEKPTIISHDRDLYGGEYSYVNMV-VNTNNDIPISGKNGTYSGLIDLINDSL 1496
QY 3531 VVALNLITY 3538
Db 1497 NSNNVDIT 1504

RESULT 14
T18399
variant-specific surface protein 1 homolog 307var1 - malaria parasite (Plasmodium fal
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18399
R:Rubio, J.P.; Thompson, J.K.; Cowman, A.F.
EMBO J. 15, 4069-4077, 1996
A:Title: The var genes of plasmodium falciparum are located in the subtelomeric regio
A:Reference number: Z18927; MUID:96324414; PMID:8670911
A:Accession: T18399
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-2042 <RUB>
A:Cross-references: EMBL:U53324; NID:g1297090; PID:g1297091; PIDN:AAB09769.1
C:Genetics:
A:Map position: 12
A:Note: 3D7vari1

Query Match 7.9% Score 1523.5; DB 2; Length 2042;
Best Local Similarity 19.3%; Pred. No. 3.7e-63;
Matches 680; Conservative 282; Mismatches 661; Indels 1895; Gaps 108;

QY 166 HDLGNVLTAKYGESEIYNNH-PHK-----GTSDACTALARSFADIGDIVRIGIDMEKPN 219
DB 12 HNLIDVLCIAOYEGSISQNHGKHOLDSTPDSQCLTELARSPALIGDVRDRDLXGN 71
QY 220 --VHREVEGLREVEFKIHDMG---EDEVKNDYNPDGSGNYKLEAMNNVNRNKEWEA 273
DB 72 NRENDLEKLEKLYFKKIDYDNLVEKKKEAEETDYK--DDAPDFYQLREDMVALNRDYMKA 130
QY 274 ITCDASVKGSGYFMQSSNPLFSPKCGHKQGVPTNLDVPOYLAWPDMGESEFCRKN 333
DB 131 ITCDAA-HDSYRRMGADGSTTESAMRQCRNAVADVPNFEDVPQYLKWFEMADFCRRK 189
QY 334 IKLKRYKDSGRND---KERLYCSHNGHDTTITMKGILHLDNKCSTCKCFEYVWL 389
DB 190 HKKDALIQCRGODGTGKDR-YCDLNRDCKRTISAKHELVOGECKKCSVVCIPFGPW 248
QY 390 GNOQEAFFKQKEREKEIQ-----SYLSDNKEFVNNINSEYK---QFEYKLETOY 438
DB 249 DNKOFEKQKKNYKNEINKKHDETTKEISGNRRKRSLLTKKMYKDEFEYKIFKD-EX 307
QY 439 ATNDFPLNLLNBEKVCYGK---GLPGSKDITPTNSADKGIFFYSEYCOVCPDGY----- 490
DB 308 PDVDAKFLDLSTKSTACESEPYPDEPRTISINEFK-YKNPDIFFSTHEYCOACPMCGMTYLL 366
QY 491 ----KCDGIKTYKSDNDRERNEDYKRPWGKPTNITVLYSGNBOGDITOLEFCN 545
DB 367 MVAWGTQNRVATIKLFKKNTLIQN-----TTDIPILPDTKSLIVKYNFCN 416
QY 546 SSNTYADKNQKQECYKXENIRCKLEONTEYNNDN-----PKTISFNHFE 593
DB 417 SSDNNSDQINMGOCHYD-----ESKSGQDNQCVGWTQNFQKQKVTSYNAFW 467
QY 594 LWTYLLRPTIKNDKLTCTCINNTTHCIDEARNCLCPDRWKQKKEEENSILKILFTK 653
DB 468 KWSSEMLDQIKWRAEDLCKLMDKKTGKKNYKNCILNFKLGNLKKKNGKQKHFNK 527
QY 654 KNIOQSYYSS---NINLEGEYFFKYMDKLKDE---AKWKELENIKKRKNFSNLENN 706
DB 528 KKIDRLVCLLKHVLLVLDVFLQDMKRAQSPQHAKIKELKKNDKDEVNNSNMTI 587
QY 707 RQYLENAIELLDHLKETATICKDNNTNEACETSHNATTNFCVKKPRGCTQPTKNIKEIAQ 766
DB 588 FDR-----LQEEBQDQKCVSNNP-EKCEETQKRPDTGA--PGA----- 625
QY 767 YFRSAVEARNRGLKHLKGAHEGIYKRGGRKDFKDNLCRIMIKHSNNLGSNPGD 826
DB 626 -----GSP 629
QY 827 GKGTGDIOTRFVYGTWEVDPBHRKHEDVIMPRRRHICTSNLEHLOTDDHPLNGNI 886
DB 630 DTGTDNLE-----DIDSD--GEEDDV-----SHVDEEPEDNPVGE-- 665
QY 887 VDDLVNNSFLGDVLLSAKYEANKIIMYKKEKNLKGPKVEYTDKHOITJIRAIRYSADI 946
DB 666 -----SSEKEQEVK---DTEAAYPKDQTPKEVNPCKIYEELF--- 703
QY 947 GDIIIRGLMERGDWKLQGHLETFGNHKLKSGKNDKYNDADAPKYLKLENNMEAN 1006
DB 704 -----KSTKNFEA- 712
QY 1007 RAKVWEMKCDIYLLKDKSGHSTQSSYCGYSDHTPLDXYIPQKLRMTWMAEMWYKQV 1066
DB 713 -----CGLKYGKNYG-----MKC----- 725

QY 1067 KEYDKLKECKECKDKDNGQCGCTKESGTGCTKCTEACNEYNIDITIGLMEQMIISDKYK 1126
DB 726 ----- 725
QY 1127 LHEQAQMSVNSNGIEASSTAKNHIDRNVLEFLSELVQONGKSNKSGTSDSAVIGNTT 1186
DB 726 VHH-----TSDKSEPTLGHSHVANS----- 747
QY 1187 YENWGAYLHDTGNFDQSQONEFCDEKSDGKDNKEKAYRDKPODHDGACGSKSPTRY 1246
DB 748 -----ADAPSGDK----- 756
QY 1247 QIKTKKAEEKTECKTVNDILKENDKKOVEDCHPKKNSNGYPMOCGINILVEDPRVC 1306
DB 757 -----DGAIC 761
QY 1307 MPRRQKLCVHFLANDNEIKLQSOYNLKAFTKSAAEFFESWYK----- 1354
DB 762 IPRRRRLYLHKT---EGVDTTDDKSLRKWFTESAVERFELMDRKKLITPOSGSPLL 817
QY 1355 -----SKDEGNELDKELKEGKIPPAFRSMFEYTGDRDFLEGTDISKHG--- 1401
DB 818 GGMITGVGVNDDENNPP-EKLLQGEIPDGFILRQHEFTIGDYRDLIFSGDKKKNQYSD 876
QY 1402 -----ESKLEQIDSLFKNGDQ-----KSPNGKTRQEWNTSHETHEAMLC 1444
DB 877 IVSGDNYIKERENTIKERIASFPONGKREGTPHYKPN-VQTPQWKNDAKHIMGMIC 935
QY 1445 ALKIKAKNDPFTENGYNNVKSFSKSTILEFAKROFLRWLIEWDDCYRQYLYKD 1504
DB 936 ALN----- 938
QY 1505 VOECKSNQDLKCDTECKNKCEDYVYKMKKKEMIPDQKYYKDERKRPEDROHIGMYT 1564
DB 939 YEETSGSDEK-----KLEKDAYKKR----- 962
QY 1565 DYGTNATDYLANRFTASCDKPGSASVORNIQLEKQAYYADAKHCGCTKTEINDKY 1624
DB 963 --GTGPNPLPQGTNGTSNEPIS-----QYQYD----- 990
QY 1625 TNISDKCKGLKVEANTGAIKQONKGPN---NYNNLKELETEVLPSPRRRLICFHALDG 1681
DB 991 -----QVYLKEENNGMASTSPSTSAABSDNPTTLQVLRPT----- 1029
QY 1682 NYTDPEYKDENGLRKRLMEVAAETGYNLQGYKEKEKEKIKTSDAHKYSYEVPCSAM 1741
DB 1030 ----- 1029
QY 1742 YSFYDLRDIILGIDNLEDEKQKTEENLKIFKNNGTISVGKSDSTYGNPGSTARKFFWNE 1801
DB 1030 ----- 1029
QY 1802 NKQCVNNAMITGYKRGHDQNSGNSARSDEDLKCKGSVPSEDDYPMCKNDEGTAVQFLR 1861
DB 1030 -----YFR 1032
QY 1862 WFAWGEDFCRKEKELEKLVGAC--NDYTC-GDNED--KRRK-----CTDA 1903
DB 1033 YLEBMOGNFCRKEKRLRLAQYEDCRGNDKYCSGDGCEVRRQDSKISFNPCGCGRE 1992
QY 1904 CIOYKFFISEMKPOYEQIKKYGKNDKIYSEHPVAKADADAREYLDQKLKICENKSGD 1963
DB 1093 CRKYKWKIKTRTEFEQKKAAYDRDK----- 1120
QY 1964 CEYKCMKDVSTQRLTDONSQNPASLDDEPKVEVGKCNQVPRGPPRVARETPSPVSLI 2023
DB 1121 ----- 1120
QY 2024 SKATASKKEAKTAPPTKOPKVENLFTEMRAQTRTRRAAQTRKRSTATTESDVGTW 2083
DB 1121 ----- 1120


```

Db 296 EYPTVYDYVQYLRMEFEAEDEFCRRKKRIENAIKNGCEKNER-YCDLNGNCEETA 354
QY 364 MKKGIHLNDKCTCSTCKVEEVLGNQOEAFKOKEREYEQ---SYLSNDKRYN 419
Db 355 RGAEIFVKGDDHCSCVACDRFYKIDNOKEDEKOKKKYDEIHGTITGKIN 414
QY 420 NINSEYKOPFELKATOTATNDTFLNLNECKYK--GGLGEKITTNSADKGY 477
Db 415 NL--VGFHFYKILKK-YYPYDKSLQKLNDEAIKPPNVEKSTDFNNEVTTFS 470
QY 478 RSEYCOVPCDGYKOD---GKRYTHKS--DNDRERNNDYKPRMGVKNITVLSGNE 532
Db 471 HTTYGACWCMGAKOKKNGGKAKKSCAKKERIFNKE-----NSTDIKILPEKG 523
QY 533 OGDITOKLEFNCSSNTYKDNKNQWECYKDEINRCXLEONTLEIND----- 581
Db 524 RSKTLEKTFCKDGOKIK---NDIKCHYDNG-----TDQDTDSNOCVLGDGMLRK 575
QY 582 NKRIISFHFELMYTYLLRDTIKMNDKLTCTN--NTTHCIDECRNCLEPDRYKOK 640
Db 576 EDKIMSYNAFEMWVMDLIDISIKWDEHGRICNKDKGTCIKGCKKICQKWEOKK 635
QY 641 EEMNSIKKLETKKNIQSYYSNINLFGYEFKYVDKLDKDEAKKELMENTIKRKNFE 700
Db 636 TEMGKIKDHFRRKOKDIPKDW----- 655
QY 701 SMLENNRDYLENAI--ELLDLKETAITICKDNTEACETSHNATNCPVPRGTOPT 758
Db 656 ---THDDFLQTLMLKDLLEITIDTY-----GDANE----- 683
QY 759 KNIKELAOYFKNSAYEABANRGLKLGKAHEBITYRGRRRDFKNLCRIMIKHSNRL 818
Db 684 --IKRI---EALLEQA--GVGGIDFALALGYKGFVAE--KDTTIDKLQHOOK-- 729
QY 819 GFSNGCDGKGTGDIQIOTRFVVTGEMVEDPBEHRKDEHYIMPRRHHICTSNLEHLOT 878
Db 730 ---BADCKLTHTDTCPPQ-----E 747
QY 879 DHPDNGNIYDVLNNSFLDVLISAKYEANKIIRMYKERNNLKGPVEYDPKHOTTICRA 938
Db 748 DRSV-----ARSESATV-----PSPPADPK----- 767
QY 939 IKSFPADIGDIIRGRDLBERNGDMVLOGLHLEVPGNINHSKLGKGNKRYNDAPRYLKL 998
Db 768 ---ATEEVANASSDEDEDFEEB----- 787
QY 999 RENWMEANRAKYWEAMKCDIKYLKDKSGHSTQSYCGYSDHPLDYIPIQKLRMTENA 1058
Db 788 --- 787
QY 1059 EWCYKQKEKYDKLEKCECKDKDKNGOGCTKESGTCTKTEACNEYNDIIGLMEOWN 1118
Db 788 ---EEEBEDEGE----- 796
QY 1119 IISDKYKELHEQAOMSVNSGIEASTAKNHIDRWITEFLSELYOONGKSNKSGTSDS 1178
Db 797 ---EEAE----- 800
QY 1179 AVIGTMYTENGAYLHDTGNFDQSONEFCEDESKDGKNEKYAFRDRPOJHDGACGK 1238
Db 801 ---EVOBEKTDSESTEAVA----- 816
QY 1239 SGSKPTRVQIKTKKAEKEDTE--CKTVNDILKENDGKOVEDCHPRKNSNGYPMQCGNI 1297
Db 817 -PSPPTQTDGVPAQEDDVKCSIVDKALK--GKLD-DACLTLYGKTAPSMKC-- 868
QY 1298 NLVEDR-----VCMPPRQKLCYHFL----- 1319
Db 869 -IPSDTKSVATTGSDTGGSGSICVPRRRRLYGLKLDWAGETTEAKSOETSGOKTFS 927
QY 1320 ANDN-----EIKLOSVMLEAFIKSAAEFFSYYYSK----- 1356
Db 928 GNSSPSEKLPQGPTEPTEJTKETPESSLHAFVSPRLRRLPWHKREQWKAQHAGAGATG 987

```

```

QY 1357 -----DGEENLDEKLEKGIKIPALIRSMFYTGDRDLEFGTD-----ISK 1398
Db 988 LQLPGVYDSDPDPOQTQKRGNIIPNDIFROMEYTLGDRDIDICIGDRDIDVGTIVSITE 1047
QY 1399 GHGEGSKLEQIDSLFKRNDOKSPNCK-----TROEWTEHSHETWA 1441
Db 1048 GESTKKIKISKILIEGLKQYVTSPPSPRDTSSRTVPHPQTSVEKTPPOOTWEANGPHIMNG 1107
QY 1442 MLCALV-----KI-----GAK-----KDDPTENYGNVNFSDK 1470
Db 1108 MICALTYEDSGAIGOPPOKVEDADKLEKLPNTANGIKMYLKEDWTSSAMPSSSSSG 1167
QY 1471 ST-----TLEEFKAPPOELRMLTEWYDYCYTRORYLADVOEKRSND----- 1513
Db 1168 SNDPINTPLTEVEIETFEFRYLHEGQNFCEKRMRLQIYNECKYVGENGYGRGKOKT 1227
QY 1514 -QLKC-----DT-----EONKCEYVYKMKKK--EMLPQDK----- 1543
Db 1228 PQSCYGEDECDQLSKYSYDYVADLECPKCAKRCMYKWKIEKKKDETEQEKAPFKOD 1287
QY 1544 YKDERDKRRFRD-----OHIGVNTDYTGATDYL-----NRKFTAS 1582
Db 1288 YVYNGNKKGGGNGFCITILKSLSDAQFLEKLGCSCKKDSNGDKLNFSPNETFVA 1347
QY 1583 CGDKPGSASVONIQLEKQAYYADKHCQCTFLIENDKYTNISSKDKGLVKEANT 1642
Db 1348 TNCKPCSEFKIDCKENGKCKNGGCTENCGTFT-----TSENFKOKG--QTAKE 1397
QY 1643 GAIKMONKGPNNNNIKELTEVY-LFPSRL-----RICFHAL-----DGWYTPPEVK 1689
Db 1398 FVMRVSDNNPNNGDILNEACQNAIGIPKSIKDKDEWEGKCGYEVCIPEKONGYTTSGENN 1457
QY 1690 DE---NGLRKLMEVAATEGYNLQYKREKKEKEXIKTSDAH-KYSYEVPPC----- 1737
Db 1458 DOIITIRGLVAHVQ--NFLDDVN-----KIKKIKISHCKNSSGTYCTIKKCVQD 1504
QY 1738 --SAMKSYFYDLIDILIGIDNLEDEKQKTEENLKI-----PNKN--GTSVGKG 1782
Db 1505 WISTKRTEWNTNIK--ILLN---EQYKNDPNYVNTKITLODLQSIDENKAIKPGCT-LTKF 1558
QY 1783 SDSTGNPBGSTARKFFWNNKEKCEVWNAAMICGYRGR---DDGSGSASDEBLKRC--G 1837
Db 1559 EDSGGINGAESSEKKNHGE---YDAIDOMLNRLODKIDDCRKNHAONGENGQACERH 1613
QY 1838 SVPSDDDYPMGKRDGTAYQFLRMFAEWGDEFC-KHKEKELEKLVGACNDYCGDNEDK 1896
Db 1614 SAPDEDDDEAIEENPYTO-----PNICPAPPEPKAEKGG-----CEFAEKK 1656
QY 1897 RKKCTDAGTOYKKKFISEMKPOYEKOIKKYGENKOKIYSEHPV--AKDADAREYLDKOLK 1954
Db 1657 EKV-----EKEKEKYVNTVAKPTEKEAAGDPAPADSEENPE----- 1694
QY 1955 KICENKSGDCEYKCMADYVSTQRLTDGNSQNMPS 1988
Db 1695 ---EERAPPEVEYTKRDKAPVKPT-----PAS 1718

```

Search completed: April 28, 2003, 10:32:38
Job time : 211.847 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:46 ; Search time 108.055 Seconds

(without alignments)
4367.892 Million cell updates/sec

Title: US-10-087-013-2

Perfect score: 19407

Sequence: 1 MFSCKRPIIKMGNAASLE.....IDLNDLSVLNLIYMKY 3542

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

al number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: A.Geneseq.101002.*
2: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
9: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
10: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19407	100.0	3542	22	AA62142
2	3489	18.0	3060	18	AAW22475
3	3489	18.0	3060	21	AAV77905
4	3108.5	16.0	2913	17	AAW00384
5	2972.5	15.3	2710	18	AAW22482
6	2972.5	15.3	2710	21	AAV77904
7	2972	15.3	2703	16	AAW70236
8	1749	9.0	2197	21	AAW18352
9	1743	9.0	2182	18	AAW22476
10	1743	9.0	2182	21	AAV77906

11	1681.5	8.7	2228	20	AAW93944	P. falciparum PFBM
12	1244.5	6.4	1726	17	AAW00385	Truncated Plasmodi
13	1045.5	5.4	1700	21	AAW18144	Plasmodium falcipa
14	1032.5	5.3	700	16	AAW70235	P. falciparum EBL-
15	1032.5	5.3	700	18	AAW22481	Plasmodium ebl-2.
16	1032.5	5.3	700	21	AAV77903	P. falciparum ebl-
17	827	4.3	407	21	AAW62148	P. falciparum VARC
18	811.5	4.2	431	22	AAW18350	Plasmodium falcipa
19	665.5	3.4	1086	23	AAW76760	Plasmodium falcipa
20	655.5	3.4	294	22	AAW62147	P. falciparum VARC
21	636.5	3.3	1604	16	AAW70105	TNF-R-EBA 175 fusi
22	633	3.3	921	18	AAW22480	Plasmodium E31a.
23	633	3.3	921	21	AAV77902	P. falciparum ebl-
24	631.5	3.3	786	16	AAW70234	P. falciparum E31a
25	631.5	3.3	1763	16	AAW41043	CD4-EBAl75 fusion
26	617	3.2	1435	16	AAW70232	P. falciparum SABP
27	617	3.2	1435	18	AAW22477	Silastic acid bindin
28	617	3.2	1435	21	AAV77900	P. falciparum SABP
29	609	3.1	311	22	AAW62150	P. falciparum VARC
30	599.5	3.1	1421	23	AAW76764	Plasmodium falcipa
31	594.5	3.1	1501	23	AAU76762	Plasmodium falcipa
32	591.5	3.0	1143	23	AAU76759	Plasmodium falcipa
33	572	2.9	308	22	AAW62151	P. falciparum VARC
34	550.5	2.8	351	22	AAW62149	P. falciparum VARC
35	529	2.7	616	23	AAW50533	Undenitified amino
36	523.5	2.7	440	21	AAW18146	Plasmodium falcipa
37	495.5	2.6	445	22	AAW66344	Malarial parasite
38	431	2.2	749	16	AAW70233	P. falciparum EBL-
39	431	2.2	749	18	AAW22479	Plasmodium ebl-1.
40	431	2.2	749	21	AAV77901	P. falciparum ebl-
41	421.5	2.2	1979	21	AAW18171	Plasmodium falcipa
42	398.5	2.1	5024	22	AAW82935	S. epidermidis ope
43	396	2.0	10182	23	ABP38314	Staphylococcus epi
44	384.5	2.0	2573	21	AAW18234	Plasmodium falcipa
45	382.5	2.0	972	23	AAU76761	Plasmodium falcipa

ALIGNMENTS

RESULT 1
ID AAB62142 standard; Protein; 3542 AA.
XX AAB62142;
AC AAB62142;
XX 29-MAY-2001 (first entry)
XX P. falciparum FCR3, varCSA protein.
XX DE FCR3, varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
XX KW erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
XX KM malaria; protozoacide.
XX OS Plasmodium falciparum.
XX PN WO200116326-A2.
XX PD 08-MAR-2001.
XX PF 01-SEP-2000; 2000MO-US24195.
XX PR 01-SEP-1999; 99US-0152023.
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
XX Gysin J, Pouvelle B, Fujii N, Smith J;
XX WPI: 2001-235109/24.
XX N-PSDB; AAF57301.
XX Novel FCR3, varCSA protein, useful for modulating parasitized red blood

PT cell binding, sequestration and onset of maternal malaria -
 XX
 PS Claim 12; Page 63-71; 78pp; English.
 CC The invention relates to a *P. falciparum* FC33, varCSA protein, that is
 CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
 CC corresponding *P. falciparum* erythrocyte membrane protein 1 (PEEMP1)
 CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
 CC protein and the encoding gene are useful for treating and preventing
 CC maternal malaria in a patient identified at a risk for contracting
 CC maternal malaria or in a patient afflicted with maternal malaria. The
 CC present sequence represents the *P. falciparum* FC33, varCSA protein.
 XX
 SQ Sequence 3542 AA;

Query Match 100.0%; Score 19407; DB 22; Length 3542;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3542; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MGFSCKYFLIKGNMASSLSGDAKSPITIKSHKSNVLEKAKNIRHPSKAKEHVDL 60
 1 MGFSCKYFLIKGNMASSLSGDAKSPITIKSHKSNVLEKAKNIRHPSKAKEHVDL 60
 61 KGDITKAEPFGSPFPVKNHNYYPYPCNIDHKEHTNLRYDDVNLHPCHGEONFEDD 120
 61 KGDITKAEPFGSPFPVKNHNYYPYPCNIDHKEHTNLRYDDVNLHPCHGEONFEDD 120
 121 ESEEGCNKIKNTKRNDAIDVAFBSRINCDKNLEALNDINTONIHDLGNYLVAKYEG 180
 121 ESEEGCNKIKNTKRNDAIDVAFBSRINCDKNLEALNDINTONIHDLGNYLVAKYEG 180
 121 ESEEGCNKIKNTKRNDAIDVAFBSRINCDKNLEALNDINTONIHDLGNYLVAKYEG 180
 181 ESIYNNHPIKTSIDACTALARSFADIGDLYRGIDMEKPRVHOKVEGIRFVYKAKIHDOGE 240
 181 ESIYNNHPIKTSIDACTALARSFADIGDLYRGIDMEKPRVHOKVEGIRFVYKAKIHDOGE 240
 241 DEYKNDYNPDSGNYKYLEEAWNNVNNKYMALITODASIKSGYMOSESNPFLSPNPK 300
 241 DEYKNDYNPDSGNYKYLEEAWNNVNNKYMALITODASIKSGYMOSESNPFLSPNPK 300
 301 GHKOKRVPMDYPOYLYKMPDEMGEEFCRKNITLKYKDSGRNDKERITCSANGHDT 360
 301 GHKOKRVPMDYPOYLYKMPDEMGEEFCRKNITLKYKDSGRNDKERITCSANGHDT 360
 301 GHKOKRVPMDYPOYLYKMPDEMGEEFCRKNITLKYKDSGRNDKERITCSANGHDT 360
 361 TTIMKKGILHLNCKTDCSTKCVFVWLVGNQOEAFFKOKKEKEKEIYSIYNDNRKFPVNN 420
 361 TTIMKKGILHLNCKTDCSTKCVFVWLVGNQOEAFFKOKKEKEKEIYSIYNDNRKFPVNN 420
 421 INSEYKQFYELKETOYATNDTFLNLNEGKYCKGLPEGEKDTITFNSADDKGIFYRSE 480
 421 INSEYKQFYELKETOYATNDTFLNLNEGKYCKGLPEGEKDTITFNSADDKGIFYRSE 480
 421 INSEYKQFYELKETOYATNDTFLNLNEGKYCKGLPEGEKDTITFNSADDKGIFYRSE 480
 481 YQVCPDGVKCDGKIKYTHKSDNREERYNNNDYKRPNGVYKPTNTVLYSGNEOGDITOKL 540
 481 YQVCPDGVKCDGKIKYTHKSDNREERYNNNDYKRPNGVYKPTNTVLYSGNEOGDITOKL 540
 541 ENFCNSSTNYKDKNNQKWECEYKDEINIRKCLEONTEINNDPKIISFHNFEFLWYTL 600
 541 ENFCNSSTNYKDKNNQKWECEYKDEINIRKCLEONTEINNDPKIISFHNFEFLWYTL 600
 601 RDTIKWMDKLTCTINNTHHCIDECNRNCLCFDWRVYKOKKEEWNISIKKLTYYKKNNIOOSY 660
 601 RDTIKWMDKLTCTINNTHHCIDECNRNCLCFDWRVYKOKKEEWNISIKKLTYYKKNNIOOSY 660
 661 YSNINNLLEGYEFKVMKLDKDEAKKELMENIRKKNKESNLENNNDYLENALEILLDH 720
 661 YSNINNLLEGYEFKVMKLDKDEAKKELMENIRKKNKESNLENNNDYLENALEILLDH 720
 721 LKEPATICKDNNTVACETSINATTPCVKPRGSGTOPTKNIKETAOYFKRSAYEARNRG 780
 721 LKEPATICKDNNTVACETSINATTPCVKPRGSGTOPTKNIKETAOYFKRSAYEARNRG 780
 781 LHKTKGAHEGITYKRGRRKDFKNLCRIMTKHSNRNLGSGNCGDKGIGDGIQTFFV 840
 781 LHKTKGAHEGITYKRGRRKDFKNLCRIMTKHSNRNLGSGNCGDKGIGDGIQTFFV 840

781 LHKTKGAHEGITYKRGRRKDFKNLCRIMTKHSNRNLGSGNCGDKGIGDGIQTFFV 840
 841 GTEWEVDPEHMRKDHEDVIMPPRRHICTSNLEHLOTDDHPNLGNIVDDLVNNSFLGDLV 900
 841 GTEWEVDPEHMRKDHEDVIMPPRRHICTSNLEHLOTDDHPNLGNIVDDLVNNSFLGDLV 900
 901 LSAKYANKITIMYKKNLKGPKETDPKHOTTCIRAIRYSFADIGDIRRDLWERNNG 960
 901 LSAKYANKITIMYKKNLKGPKETDPKHOTTCIRAIRYSFADIGDIRRDLWERNNG 960
 961 DWKLOGHLETFVGNHKSILKRGKNDKYNDAPKTLKLENNWEANRAKVMKAMCDIKY 1020
 961 DWKLOGHLETFVGNHKSILKRGKNDKYNDAPKTLKLENNWEANRAKVMKAMCDIKY 1020
 1021 LKDKSGHSTOSSYSGYSDPHTLDIYIYOKLWMTWMAEWYCKVOKKEYDKLKECKECK 1080
 1021 LKDKSGHSTOSSYSGYSDPHTLDIYIYOKLWMTWMAEWYCKVOKKEYDKLKECKECK 1080
 1081 DKDNGOGCTKESGCTCTEACNEYNIIIGLAKRQWNIISDKYELHEQOMSVNSGI 1140
 1081 DKDNGOGCTKESGCTCTEACNEYNIIIGLAKRQWNIISDKYELHEQOMSVNSGI 1140
 1141 EASSTAKNHDNRNVEFLSELYOONGSKNSGTSDESAYIGTNTYENVAGAYLHDTGNF 1200
 1141 EASSTAKNHDNRNVEFLSELYOONGSKNSGTSDESAYIGTNTYENVAGAYLHDTGNF 1200
 1201 DDCGQONFCEBKSDGKNKNEYAFEDKPODDHAGCGGSSGKPRVQIKTKKAEKDETE 1260
 1201 DDCGQONFCEBKSDGKNKNEYAFEDKPODDHAGCGGSSGKPRVQIKTKKAEKDETE 1260
 1261 CTVNDILKENDGKQVBDCHPKKNSNGYPMOGCINILVBDPRVCMPPRROKLCVHFLA 1320
 1261 CTVNDILKENDGKQVBDCHPKKNSNGYPMOGCINILVBDPRVCMPPRROKLCVHFLA 1320
 1321 NDNEIKKLOSOVNLKEAFIKSAAAFPPSWYIYYSKDEGEJLDELKEGKIPATIRSM 1380
 1321 NDNEIKKLOSOVNLKEAFIKSAAAFPPSWYIYYSKDEGEJLDELKEGKIPATIRSM 1380
 1381 FYTFGVDYDFLFTGDISKSGEGSKLKEQIDSLFKRNGDOKSPNGKTOEWMTESHEIWE 1440
 1381 FYTFGVDYDFLFTGDISKSGEGSKLKEQIDSLFKRNGDOKSPNGKTOEWMTESHEIWE 1440
 1441 AMLCALYKIGAKKDFEBNTGYNNVNFSDKSTLLEEFKAKRPOFLRWLTJEWYDYCYTROK 1500
 1441 AMLCALYKIGAKKDFEBNTGYNNVNFSDKSTLLEEFKAKRPOFLRWLTJEWYDYCYTROK 1500
 1501 YLKDVOEKSKNDOLKCDTECNKCEDYVYVYMKKKKEMIPQDKYKDERKKRRDRHIG 1560
 1501 YLKDVOEKSKNDOLKCDTECNKCEDYVYVYMKKKKEMIPQDKYKDERKKRRDRHIG 1560
 1561 VMVTDYTGTAADYLNKFTASGDKPGSASVVOYNIQLEKQAYYDADRHGCTFIEN 1620
 1561 VMVTDYTGTAADYLNKFTASGDKPGSASVVOYNIQLEKQAYYDADRHGCTFIEN 1620
 1621 DDKYTNISDKDKGLYKENTGAIKQKNGPNYNNLKLJEDVJPPSRRLICFHALD 1680
 1621 DDKYTNISDKDKGLYKENTGAIKQKNGPNYNNLKLJEDVJPPSRRLICFHALD 1680
 1681 GNTDPEVNDENGIRKRLMVAATTEGYNLGQYKKEKKEKIKTSPAHKYSYVPPCSAM 1740
 1681 GNTDPEVNDENGIRKRLMVAATTEGYNLGQYKKEKKEKIKTSPAHKYSYVPPCSAM 1740
 1741 KYSEFYDLRDIILIDLEDEKQTEELKKEIFKNKNGTSVAGKSDSTGNGSGSTARFFVNN 1800
 1741 KYSEFYDLRDIILIDLEDEKQTEELKKEIFKNKNGTSVAGKSDSTGNGSGSTARFFVNN 1800
 1801 ENKECYVNMAMICGKRGDRDGSNGSARSDEDIKKGVSVPDDYPMGNRREGIAYOFL 1860
 1801 ENKECYVNMAMICGKRGDRDGSNGSARSDEDIKKGVSVPDDYPMGNRREGIAYOFL 1860
 1861 RMEAWGEDEFCRHKERKELELVAGCNDYTCGDNEDKRRKCTDACTQYKRFISEMKQYER 1920
 1861 RMEAWGEDEFCRHKERKELELVAGCNDYTCGDNEDKRRKCTDACTQYKRFISEMKQYER 1920

QY 1921 QIKKYGENDKITYSEHPVAKDADADAREYLDKOLKICENKSGCEYCKAMDVSTOURLTDC 1980
 Db 1921 QIKKYGENDKITYSEHPVAKDADADAREYLDKOLKICENKSGCEYCKAMDVSTOURLTDC 1980
 QY 1981 NSONMPASLDDEPEKEGKCNCOVPRGPVPRVREPSPRSLISKATASKAKKTAAPTR 2040
 Db 1981 NSONMPASLDDEPEKEGKCNCOVPRGPVPRVREPSPRSLISKATASKAKKTAAPTR 2040
 QY 2041 OPRKVENLTTEMAQOTRTARRAQQOTRRRTSTATTTESDVGTWVKAILSNKPSDRSGIEGC 2100
 Db 2041 OPRKVENLTTEMAQOTRTARRAQQOTRRRTSTATTTESDVGTWVKAILSNKPSDRSGIEGC 2100
 QY 2101 NPKTYGYRFGGCTVGSKEBENGICMPRRKKLCINNIQYLYNETENKRDNDIKEAFIK 2160
 Db 2101 NPKTYGYRFGGCTVGSKEBENGICMPRRKKLCINNIQYLYNETENKRDNDIKEAFIK 2160
 QY 2161 CAIEFOFLMKYLTLENPAENELONGTIDDEPKRIMYTYGYKMPFGTDSINCKIT 2220
 Db 2161 CAIEFOFLMKYLTLENPAENELONGTIDDEPKRIMYTYGYKMPFGTDSINCKIT 2220
 QY 2221 TYTNSVTLTENNNKKQDKKKDEBLRKIFWEKNKKFIEGMYGLTYHLTDENEKIKR 2280
 Db 2221 TYTNSVTLTENNNKKQDKKKDEBLRKIFWEKNKKFIEGMYGLTYHLTDENEKIKR 2280
 QY 2281 DNYQYDMTKLPSLEBYKRPOLFWETEMAEFCNKKREOLKLEAGCKEYECNSND 2340
 Db 2281 DNYQYDMTKLPSLEBYKRPOLFWETEMAEFCNKKREOLKLEAGCKEYECNSND 2340
 QY 2341 GKTQCEACAWYQNFITKMKTEYERQREKFKDKGKRYKDYPSERDIEKATCAHEYL 2400
 Db 2341 GKTQCEACAWYQNFITKMKTEYERQREKFKDKGKRYKDYPSERDIEKATCAHEYL 2400
 QY 2401 NMLKELCGNCKDCSCMQKPSQLPKTTQOSSDANDMPEESLDYVEEFNKCBPELSK 2460
 Db 2401 NMLKELCGNCKDCSCMQKPSQLPKTTQOSSDANDMPEESLDYVEEFNKCBPELSK 2460
 QY 2461 GSMIHTKKTITEKIPMNCVEKAAAYLSKAEENMDITLKEKPIPIESTKEKSKNMTNN 2520
 Db 2461 GSMIHTKKTITEKIPMNCVEKAAAYLSKAEENMDITLKEKPIPIESTKEKSKNMTNN 2520
 QY 2521 NPODPKPYAPDKYIGRRNRCENRENRFRKYDEMKCYNSKRYOKKKRYCVPRRHHNC 2580
 Db 2521 NPODPKPYAPDKYIGRRNRCENRENRFRKYDEMKCYNSKRYOKKKRYCVPRRHHNC 2580
 QY 2581 LRLNDEIKTERLKSNYLLKMYRTARNEGIDILKFNSENGCAMPICDTMKYSPADLG 2640
 Db 2581 LRLNDEIKTERLKSNYLLKMYRTARNEGIDILKFNSENGCAMPICDTMKYSPADLG 2640
 QY 2641 DIVRGDMRLRIGGYLPVVEIKLYVEEYLYGKMRKNKNGKNDYQTRSAMWANDARKD 2700
 Db 2641 DIVRGDMRLRIGGYLPVVEIKLYVEEYLYGKMRKNKNGKNDYQTRSAMWANDARKD 2700
 QY 2701 IMKAMTCAPEDAKFLFRKRGMDGFERITLIDCKGCHDPPVVDYIPORRMTWEMSEY 2760
 Db 2701 IMKAMTCAPEDAKFLFRKRGMDGFERITLIDCKGCHDPPVVDYIPORRMTWEMSEY 2760
 QY 2761 CSALMELEKFRKSCDCHTSDCKNDYDENKCEQCTRCOEYKKNVLMKSLFDIOSNK 2820
 Db 2761 CSALMELEKFRKSCDCHTSDCKNDYDENKCEQCTRCOEYKKNVLMKSLFDIOSNK 2820
 QY 2821 YKELYEOPITYTKISTYDHQONPVQKLTFRKSECSVESFSFSLHETSCLNRYKNEENDGS 2880
 Db 2821 YKELYEOPITYTKISTYDHQONPVQKLTFRKSECSVESFSFSLHETSCLNRYKNEENDGS 2880
 QY 2881 NIKTYAFEEETPKSYKACSCTLPSKNPLDNCPTDQNKDGCKELQOTFCSKNDYDNMLN 2940
 Db 2881 NIKTYAFEEETPKSYKACSCTLPSKNPLDNCPTDQNKDGCKELQOTFCSKNDYDNMLN 2940
 QY 2941 WNAVYVLNSSDNNKGVLLPERRRHLCITRTAYNRYKGGKEILKKLLTSAFSOGQLLQ 3000
 Db 2941 WNAVYVLNSSDNNKGVLLPERRRHLCITRTAYNRYKGGKEILKKLLTSAFSOGQLLQ 3000

QY 3001 KYKSEELCEFAKYSYADYSDIILKGTDMMDTSLSEKIKKIFETSSEATENKRTWENNR 3060
 Db 3001 KYKSEELCEFAKYSYADYSDIILKGTDMMDTSLSEKIKKIFETSSEATENKRTWENNR 3060
 QY 3061 ROIWHMLGCKYKATSVTLDEGNCOLPKDEEONFRLMILWAKOACREKHYVSLKT 3120
 Db 3061 ROIWHMLGCKYKATSVTLDEGNCOLPKDEEONFRLMILWAKOACREKHYVSLKT 3120
 QY 3121 KCPRSNEDNEEASELLROPQCONDIRKYISLNLILKNTMENLIRYKQLOKDOSSGNIDNK 3180
 Db 3121 KCPRSNEDNEEASELLROPQCONDIRKYISLNLILKNTMENLIRYKQLOKDOSSGNIDNK 3180
 QY 3181 PSEENQSYIKSKDOSCALELNDINEIVGTCKNNENNEKEVILKLYPGLYVEDETHKN 3240
 Db 3181 PSEENQSYIKSKDOSCALELNDINEIVGTCKNNENNEKEVILKLYPGLYVEDETHKN 3240
 QY 3241 HVLDSNKEEEOYVRKALYEFTHYDSFOAFLFTHRAOYDPRNDLKSSISVYVS 3300
 Db 3241 HVLDSNKEEEOYVRKALYEFTHYDSFOAFLFTHRAOYDPRNDLKSSISVYVS 3300
 QY 3301 ALGLIALHFMAKKFKSSVDLRLINIPQGEYGMPTLESKNRYIPYSGPYKGTYYIMEG 3360
 Db 3301 ALGLIALHFMAKKFKSSVDLRLINIPQGEYGMPTLESKNRYIPYSGPYKGTYYIMEG 3360
 QY 3361 DTSGDEDKYMDLSSDITSSSEYEBELINDIYVPSGPKYKTLIEVYLEPSKRDIPSD 3420
 Db 3361 DTSGDEDKYMDLSSDITSSSEYEBELINDIYVPSGPKYKTLIEVYLEPSKRDIPSD 3420
 QY 3421 TPSENDTPRTNRFIDDEENELKHPFSOYLPTBPNNNYKADIPMTEPMTLYSDNEEK 3480
 Db 3421 TPSENDTPRTNRFIDDEENELKHPFSOYLPTBPNNNYKADIPMTEPMTLYSDNEEK 3480
 QY 3481 PFIIISHRDLYTGKEISYNNKNTNNDIPNNARNDYRGIDLINDSLVNLILYMM 3540
 Db 3481 PFIIISHRDLYTGKEISYNNKNTNNDIPNNARNDYRGIDLINDSLVNLILYMM 3540
 QY 3541 KY 3542
 Db 3541 KY 3542

RESULT 2
 AAW22475
 ID AAW22475 standard; Protein; 3060 AA.
 XX
 AC AAW22475;
 XX
 DT 12-SEP-1997 (first entry)
 XX
 DE Plasmodium var-7.
 XX
 KW DBL gene family; SABP; static acid binding protein; vaccine; therapy;
 KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KW Plasmodium.
 XX
 OS Plasmodium vivax.
 OS Plasmodium falciparum.
 OS
 PN W09640766-A2.
 XX
 PD 19-DEC-1996.
 XX
 PF 07-JUN-1996; 96WO-US09508.
 XX
 PR 07-JUN-1995; 95US-0487826.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 XX
 DR WPI; 1997-052231/05.

DR N-PSDB; AA172882.

PT New malaria vaccines - contains cysteine-rich DBL family protein
PT binding domains homologous domains of the Duffy and sialic acid
PS binding proteins

XX Claim 8; Page 61-67; 96pp; English.

CC This sequence represents var-7 of Plasmodium. Var-7 belongs to
CC the Duffy binding like (DBL) family of genes which have homology to
CC Duffy antigen binding protein (DABP) and sialic acid binding protein
CC (SABP) conserved regions (see AA172889 and AA172888 respectively). The
CC var family of genes modulate cytoadherence and antigenic variation. Of
CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC protein (DABP) are soluble proteins that appear in the culture
CC supernatant after infected erythrocytes release merozoites. DABP and
CC SABP mediate the binding of merozoites and schizonts to the erythrocyte
CC surface. These proteins are necessary for erythrocyte invasion by the
CC parasite. The compositions can be used in the compositions of the
CC invention. The compositions are for the treatment and prevention of
CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
CC family of genes having homology with conserved regions of DABP and SABP.
CC The compositions are used for the treatment and prevention of malaria.
CC They are also used in the preparation of vaccines for inducing a
CC protective immune response in a mammal to Plasmodium merozoites
CC (especially Plasmodium falciparum or Plasmodium vivax).

SQ Sequence 3060 AA;

Query Match 18.0%; Score 3489; DB 18; Length 3060;
Best Local Similarity 26.7%; Pred. No. 2.5e-200;
Matches 1054; Conservative 472; Mismatches 934; Indels 1490; Gaps 148;

QY 9 IIKGMAASLEGDASPIKESHSARVLEKRNIRHPSK-YAKEHVSILKGLTKA 67
DB 4 LAMGKREAGGDDIED-----ESAKHMFRIKDYDKYKBEKRGKGLQRLSPA 56
QY 68 EFRGSPFPVKNHNYYPYPCNLDKHEHNLRYDVNLRHPCGHEQRNFDESESEC-G 126
DB 57 KEKNESDPQTPED-----PCDLKHRYTNV---TTNVLNCPADSDVRFSEYSGGCTH 108
QY 127 NKIRYKRRKNDIAICAPRRHRMCKNLEALNDINTONIHLLGNVLTAKYEGESYINN 186
DB 109 NKIKSOGDNGACAPYRHLVCDQNEQLEPIKITWTNHLVYCAAAFEQGSITOD 168
QY 187 HP-HKGT-----SDACTLASEADIGIVRGIDMFRKNVHD-----KVEGLREVEFKI 235
DB 169 YPKYATIGDSQICTMLANSFADIGIVKGRDLYLGNPEIKORQOLENNLKTIFGKI 228
QY 236 HD---GMEDEKNDYNDPGSGNYTKLREAMWNVNKNYWEAITCDASKSGYFMQSEST 292
DB 229 YEKLGAEARVGN--PE-----FFKLRDMWTANRETYKAITCNA-WGNITYF-HATNR 280
QY 293 PLFSNPKGHHGKVPYTLVDVYPOYLRFEDWGEFCKRNLIKRYKVSQR---NDKE 348
DB 281 GERTKGYRCNDVDPYEDVYPOYLRFEMAEDEFCRKKKKIKDYVRNCRGKDEKDO 340
QY 349 RLYOSHNGDCTTIIWKGILHLNCKTDCSTKCKVEVWVLGNQOEAFAKQKEXYEKEIQ 408
DB 341 R-YCSRMNGYDEKTRKRAIGKRLYRGKQCSCLYACNPYDWMNNOKEPDQOKKKYDEIK 399
QY 409 SY-----LSNDNKFVNINSE-YKOFYEKLKETOYATNTPTNLINLEGKYC--- 454
DB 400 KYENGASGSGRQKRDAGTTTNDYGEKKRYDELNSERYTDFLEKLSNEITCIRYK 459
QY 455 --KGLGEGEKITFTNSA-----DKGLIFYSEYQOVCPDGVKDGKILKYTH-KSND 504
DB 460 DEEGGTIDFKVNSDSGASGIVNESQGYFRSKYQCPYCGVK---KVNNGGSSNE 515
QY 505 RERVNN-----EDYKPPMGVKKPTNITVLYSGNEQDITOKLEFNC-----SS 547
DB 516 WEKNNKCKSGKRLYEPKPKEGTITIIILSGCHDIEKLNKFKCEKNGDTIINSGGSG 575

QY 548 TNYKDKNN-----QKMECYTKDENINCKLEQNTLEINNDPK-----II----- 586
DB 576 TGGSGGSGNSROBELYEEMKC-YKGEDVYKVGHDEDEDEEDYENVKNAAGGLCIKNKKNKE 634
QY 587 -----SEHNFELMWYLLIRDTIKNNDKIKCTCI-NNTTTHC-IDECNRNC 629
DB 635 EGGNTSEKPELDQTFNFFYYVAHMLKDSIHKKKILQRLQNGNRIKCGNNKCNND 694
QY 630 LCPDRVYKQKEEMNSIKKLTFRKNQI-----QSYSNINIFEGYF--- 672
DB 695 ECFKRWITQKKNDEKQIVOHF-KTONINGRGSNDTALIPRDHYLYQVNLQEFELKGD 753
QY 673 -----FKYMDKLDEKAKKELMENIKRKNESNLENRD-----YLENAITELL 718
DB 754 SEDASEKSENLDMEA-----ELHNLRIEISENDENASVGGGVTEQKIMDKIL 807
QY 719 DILKETATTC-----KDNNTNEACETSHNATNTPCVKPRGQTO-PTKNIKELAQ 766
DB 808 NFEKDEADICLETIHDEDEEKKGCGNCEIEGEMFRYNPCSGESGNKRYVLANKAYQ 867
QY 767 YFKRBAYEARRKGLAKLKKAHBSITYRGGRRRDEKDNLCRIMIKHSNRLAFSGNCPD 826
DB 868 MHHKAKTQLASRAGSALRGDISLAQFNKGRNGSTLKQICKINENYSNDRSGNGPCT 927
QY 827 GKGTGDIQTRFVVGTEWEVDPEHMRKDHEDVIMPPRRRICTSNLEHLOTDHPLANGT 886
DB 928 GK-DGDHGVYMRKISTEENSIEGKKQTSYKNVFLPREHREKTSNLELDV-----GSV 980
QY 887 V-DLVNNSFLGDVLLSAKYBANKIIRAYKERNKMLPKPEVETD---KHQTTICRAIYS 942
DB 981 TKNKASHSLHVDQALAKTDAEIKRYKQNNI-----QUTDPIQKDEAMCRAVYS 1036
QY 943 FADIGDIIRGRDILERNNDWAKYKLOGHLEYVGNTHKSLG-KGNDKYND---APRYLKL 998
DB 1037 FADIGDIIRGRDILERNNDWAKYKLOGHLEYVGNTHKSLG-KGNDKYND---APRYLKL 998
QY 999 RENWMEANRAKVPAMKCDIKYLLKDKSGHSTGSSGYSDHPTPLDITPQKIRWMEWA 1058
DB 1097 RADWMEANRHQVWAMK-----ATKILIC---PGMPVVDYITPQIRLWMEWA 1141
QY 1059 EMYCKYKKEVDKLEKCKECKDKNQOGCTKESGCTCTCTACNENYDILAKMEOWN 1118
DB 1142 EMYCKAQSQEDDKKLCACQMSKGDGK-CT-OGDVDCGCKAKCAKPKYKREIKMNEOWN 1199
QY 1119 IISDKYELHEQAQMSVNSGIEASSAKNHD---RNVEFELSLEYQONGK---SKN 1171
DB 1200 KISDKYMLYLQAKTSTNG-----RTVLGDDDPDYOQWDFLPIHKASIAARVLVKRA 1255
QY 1172 SGTSDSAVYGTNTTENVGAYLHDTGNFDCQSONEFCDEK-----SDGKNEKAYFR 1225
DB 1256 AGSPTEIIMAAPITPYSTAGIYHOEIGYGCQEOFOFCEKKGANSTSTKKNKRYTK 1315
QY 1226 DKPDHAGAGCGSGSPRTVOJIKTKKAERKTECKTYVNDILKENDGKVOEDCPKRN 1285
DB 1316 QPPEVATACDCINRSOTE---EPKKKEENVSACKIYEKILEGNGRTYVGECPKES 1371
QY 1286 SNGYPMQC-GNINLYEDPRVCAPRRKQCYAFLLANDNEIKKLOSQVLMKEAFIRSA 1344
DB 1372 ---YPMDCKNNDIDSHD-GACMPRRQKLCIYLIHESOTETIKTDNMLKAFIKTA 1427
QY 1345 ETEFSWYTKSK-DGCGNELDKELGKTIIPAFILSMFYTFGDYRDLFGTDSKGGHG 1403
DB 1428 ETELSQWYKSKNDSPAKTIDR---GLIPQFIRKSMYTFGDYRIDCLNTDISKONDY 1483
QY 1404 SKLEQIDSLFKKGDGKSPNGKTRQEMWTEHHEHTEALCALVYIAGAKKD---FTENY 1460
DB 1484 AKAKDKIGKFFKSGDSKSPGSLSRQEMWTKNPELTKMGLCALTYVYTDORRRKIKNDY 1543
QY 1461 GYNNVAFSK-STLYEFAKRPQFLRWLTLEWYDDCYTRKQYLIKDVQEKCS-NDQLKCD 1518
DB 1544 SYDKVQSONGNPSELEFAKPOFLMMIEMGEFECAEROKKENIKKACNENINSTOQN 1603

QY 1519 T---ECNKCEDYVYKMK-KKKEIMIPQDKYKDEBERKKRPFDRKHIGVMTDTYTGTNATDY 1574
 DB 1604 DAKHRCNOACRAQOEVEENKKEKESGQTNPNFKANVOCPQDPEYKKEYEKD----- 1654
 QY 1575 LNRKFTASGDRPGSASVYORNIOLLEKQAYYDADHCGCTKRENDKTYTNISKDKCK 1634
 DB 1655 -----GVDP-----IQGNETYLQK-CDNNK-GSC----- 1676
 QY 1635 GLVKEANTGAIKQONKPNPNYNNLKELTEDVLEPSSRLRLICFHALDGNVTPDEVKENG 1694
 DB 1677 ----- 1689
 QY 1695 BKRLMEVATBEGYNLQGYKKEKKEKIKTSDAHKISTEYVPCOSANKYSFYDLRLITLIG 1754
 DB 1690 -----PEGKY-----AHKYP----- 1699
 QY 1755 DNLEDEKOTEBELKIKIFNKNGTSVGKSDSTGNPGSTARKPFENENKECVANNMICGY 1814
 DB 1700 ----- 1699
 DB 1815 KRGRODNGNSARSDEDLKKGSGVPSDDYPMGKNRDEGTAYQFLRMPAEWGEDECKHR 1874
 DB 1700 ----- 1699
 QY 1875 EKELEKLVGACNDYTGCDNEDKRRKCTDQYKRFISEMKPQYKQIKKYGKKNKIYS 1934
 DB 1700 -----EKC----- 1702
 QY 1935 EHFVAKDAEDAREYLDKOLKICENKSGDCEYKMDVSTQRLTDGNSQNPASLDEBPK 1994
 DB 1703 -----DC----- 1704
 QY 1995 EWEKGCNCOVPRGPPRYRRETPSPVSLISKATASRKEAKTAPPTQPKVENLITTEMA 2054
 DB 1705 -YQK--HYPSIRP-----PPPYQPO----- 1723
 QY 2055 QRTIRAAQOTRKRTSTATTTESDVGTMKAILSNKPDNRSGLEGCNPTTYGY--PKWG 2112
 DB 1724 -----PEALPYTVVDCSIVKTLTK--DTNNEFDACGLK-YGKTAPSSWK 1764
 QY 2113 CI-----VGKSKENENGICMPRRKKLCINNIOYLANET-----ENKRD 2151
 DB 1765 CIPSDRKSAGATTGKSGSDSGSICIPRRRLRYVGKLO--EWATLPLQEGAPASHSRA 1822
 QY 2152 NDIKEAFKCAIETQFLMLKTYIE--NPAEEN-----LONGT 2188
 DB 1823 DDLRNAPFIOSAIETEFELMDRYKEEKKPOGDSQALSQLSTSYSDDEEDPDKLLONGK 1882
 QY 2189 IPDEFERINYTYGYDKDMFEFGDISDKKIIVTNSVTTLLENKKKKODKKDEELK 2248
 DB 1883 IPDEFILMEFTYLDGYRDLILVHGNTSDSGNTNGSNNNNIVLEASGNKEDMOKIOEKIEQ 1942
 QY 2249 I-----FWENKPKFIWEGMIYGLTY-----HLTDENEKEKIR 2280
 DB 1943 ILPKNGTPLVPRSSAQOTPDKNMNEHAESIMKMICALYTEKNPTTSARGDENEK 2002
 QY 2281 DMT-----QYN-DMTKL-----TPSLEEEYKRP 2302
 DB 2003 EYVEKEFGSTADKHGTASPTGTGYKQYDYEKVKLEDTSAGKTPSASDPPLISDYLRLP 2062
 QY 2303 QFLRMETNAEEFCNRKREBOLKLEAGCKEYE-----CN---GSN 2339
 DB 2063 PYRYVLEENQONCKRRKHLKQIKHECKYEEBGGSRGCGITROYSGDCEACENEMLPKN 2122
 QY 2340 DGTQOE-----CAEACVYQNFILKWKTEYEROREKFKDKD-----GKKYKDYPTSERDI 2390
 DB 2123 DGTVPDLKPSKCAKPCSSYRKWIESKGEPEKEQKAYEQDKCVCNGSNKH----- 2174
 QY 2391 EKATCAHEYILAMKLEKCKKOCSCMOKPSSQLPKTTQOSSDANDMPSLDIVBEEN 2450
 DB 2175 -----NGFCETL-----TTSKADDFLTKIG----- 2195
 QY 2451 KCCEPBLSKKSGMI-----HTKKTIEP--KIPMNCVKAAYILSKAEANMDITLKE 2500

DB 2196 --PCKPNNEGKTIFDDDKYTFKHTKD--CDPCLKFSVNC----- 2230
 QY 2501 KFIPISTKEKESKNSGTNNNPNPCDPKPKPYAPDKYIARRPDCENRENRPKVDYEMKCYKN 2560
 DB 2231 -----KKDECDNS-----KJDCRN-----KN 2247
 QY 2561 SKFYOEKRRVCPVPRREHMLCRLNDEIKIERLDSNYLLKMYRRTARNBGIDIKKNNS- 2619
 DB 2248 S-----IDATDIENGVSDFVLEMRVGSADKS-----GFND 2278
 QY 2620 --ENGCAAMPICDTMYKFSRADLGIYRGIDMLRIGSYLPEVLEIKLYVEYIYGAKRNNK 2677
 DB 2279 GLENAC-----RGAG-----IFEGI----- 2293
 QY 2678 KGRNKYNDVOTFRSAMWANDNRKDYNKA-----MTCKAPEDAKLFKRGMDPERITLQ 2731
 DB 2294 -----RDEMCKCRVCGYVYCK--PENVGEAKK-- 2321
 QY 2732 DKCGHKDPPVDYIPQFRFMTWSEYKCALMELEKFKKSDCHKTSDRCKNDYDEN 2791
 DB 2322 -----HIQIALYKRWVEYF--FEDYNNIKH--KISHRIK--G 2355
 QY 2792 KCEQCKTRQOEYKNEFLKKSLEFDIOSNKYKELYEQPIYTKISTYDHQNFVQKLTFS 2851
 DB 2356 EISPC-----INKVEKM--VDQKRKEKETER-----FKD 2385
 QY 2852 ECSVESFSEYLAHETSCLYKFNENDGSSNIRTYAEETPK-----SYKEA 2897
 DB 2386 Q-----YK-NDNSDDNVRSEFLETLIPQITDANAKKVIKLSFGNS 2426
 QY 2898 CSCTLPKPNLDCPTDQNKDGCKELQTFCSKNDYDNNLDNMAVLYLNSDDNKGVL 2957
 DB 2427 CGCSASA-----NEQKNK-- 2440
 QY 2958 IPPRRHLCRPTTAYNRYKRGDEILKLLTSAFSQGLGKYSEBELCEAKKYS 3017
 DB 2441 ----- 2440
 QY 3018 ADYSIIKGTDMMDTSLSEKIKKIFETSNEATENRKRTWMENNRQIMHAMLCYKIATSK 3077
 DB 2441 -EKKDAI--DGLKRLKDKI----- 2457
 QY 3078 VTLDEGWCQLPKDEFTNOFLRLIEMAKOACKKHHVSLSKTKC-----PRSNEDNEFA 3132
 DB 2458 -----GEC-----EKKHOTS--DECSDDTPQPTLEB----- 2483
 QY 3133 SELLRQPGQONDIRKIYISLILKNTMENINIKYKLOKQOSSGNINDNKPSEENVQSYKS 3192
 DB 2484 -EYL-----DDDIETEBAKNNMPKICEVY--LTAQOODE--GGCVPAENSEEPA-----A 2530
 QY 3193 KDSQCALELNDINEIYVGRKNNEENFEKVLKLYBGLYFVEDETHKNHVLNDINKEEQ 3252
 DB 2531 TDS-----GKETPEQTPYLK-----PEEA 2550
 QY 3253 TVRPKALYFTPHVDSFYQAPLESTHRVAYDKN-----DLKSSISVYIYSAIG 3303
 DB 2551 VPEBP-----PPPOEKAAP-----IPQPPPTPTQLDNPVHTALTAVTSTLMSVG 2599
 QY 3304 L-----IALHEMKKFKSSV--DLRLIINIPQEGFMPLSKNRYIYPRSGPYGKTYIYM 3358
 DB 2600 IGFATFTYFLKKTSSVONLFOIOLIPKSDYDIPKLSPNRYIPTTSKYGKRYIYL 2659
 QY 3359 EGDY-----SGDEKYMMDLSSDITSSSEYEELDINDIYVGSPPKXTLLLEVLLPSKRD 3415
 DB 2660 EGDGSDSGYTDHY-----SDITSSSEYEEMDINDIYVGSPPKXTLLLEVLLPSGN 2713
 QY 3416 IPSD-----DTPS-----NTPRNRKIDENMELKIDFYQYL--- 3449
 DB 2714 TTASGNNTTASGNNTTASGNNTPSQNDIONGIPSSKITTENMQLADEFISOYLQSE 2773
 QY 3450 PNTPEPNNTKASADIPNTEPNTLYSDNPEEKPLISLHBDLTGKEISY----- 3499

Db 2774 PNTBP--NMLGYNVNDNTHPTTSH-HNVEKPFMTSHDRNLFESGEENYNDMNSGNPPI 2830
 Qy 3500 NINNSTN-----TNDIPNARNDSYRGIDILNDSLVYNTLITY--MMK 3541
 Db 2831 NISDSTNMSDLSLNNHSFYNDRKNDYSGIDILNDSLVNTHIDYDEMUK 2880

RESULT 3

AAV77905
 ID AAV77905 standard; Protein; 3060 AA.

AAV77905;

13-JUN-2000 (first entry)

Plasmodium var-7 polypeptide.

DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein; DABP; Stialic Acid Binding Protein; SABB; malaria; vaccine; Immunisation; protozoicide; var-7.

Plasmodium sp.

US5993827-A.

30-NOV-1999.

07-JUN-1995; 95US-0487826.

10-SEP-1993; 93US-0119677.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Sim KL, Chittis C, Peterson DS, Su X, Wellens TE, Miller LH;

WPI; 2000-194198/17.

N-PsDB; AA298287.

Isolated protein binding domains from Plasmodium vivax and Plasmodium falciparum erythrocyte binding proteins useful for vaccinating against malaria -

Disclosure; Columns 109-124; 93pp; English.

The invention relates to ebl-1 polypeptides that are encoded by the DBL (Duffy-binding like) gene family. The ebl-1 proteins are substantially identical to the Duffy Antigen Binding Protein (DABP) and Stialic Acid Binding Protein (SABB), which are soluble proteins that appear in the culture supernatant after erythrocytes infected with malaria release merozoites. Immunochemical studies indicate that DABP and SABB are the respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy and stialic acid receptors on erythrocytes. The ebl-1 polypeptides may be used to vaccinate against malaria, especially caused by P. falciparum. Immunization with the polypeptide provides effective protection against malaria. The present sequence represents the var-7 polypeptide.

Sequence 3060 AA;

Query Match 18.0%; Score 3489; DB 21; Length 3060;

Best Local Similarity 26.7%; Pred. No. 2.5e-200;

Matches 1054; Conservative 472; Mismatches 934; Indels 1490; Gaps 148;

Qy 9 IIRGNASSLEGDAPSIPIKESHSARNVLEFARKNIRHPSK-YAREHDSLKGLDTKA 67
 Db 4 LAKGPREAAGDIDED-----ESAKIHMDRIGKQVYDKVKEAKRGKGLQGRLESA 56
 Qy 68 EFRGPGSTPVANKHNYYPYPCNDLHKETNLRVDVNIHRPCHGREONRFRDEDESEC-G 126
 Db 57 KFEENESDPQTPED-----PCDDLHKYHTNY--TTNVINPCADRSVRFSDEYGGCTH 108
 Qy 127 NKIRYRKNDATACAPRRRHMDKRLLELNDINTONIHDLGANTLVYAKYEGESTVNN 186
 Db 109 NRIDSGOGDKKACAPYRRRLVACDNLQLEIPIKTIINTNHLVLVYCAAKFEGOSITQD 168

Qy 187 HP-HKGT-----SDACTALARSFADIGIVRGIDMEFKPVND-----KVETGLAREVFKKI 235
 Db 169 YPKYQATYDSDPSQICTMLARSFADIGIVRGIDLVGNGPQIKRQOLENNLTKITFGKI 228
 Qy 236 HD---GMEDEVANDNPDGSGNYTKLRKAWNVNVRNRYWEALITCDASYKSGYFMOSESNT 292
 Db 229 YEKLNAGEARVYND--PE---FFKLREDMTANRETVWAKITCNA--NGNYTF-HATCNR 280
 Qy 293 PLESNPKCGHKQKAPYTNLDYVPOYLRFMEDEGESEFCRNRINKTKYVDSOR---NDKE 348
 Db 281 GERTKGYCRQNDQYPTFYDYVPOYLRFMEDEGESEFCRNRINKTKYVDSOR---NDKE 340
 Qy 349 RLYCSHNGHDCITTTTWKGLILHLNKCQDCSTKCVFPEWMLGNOEAFKQKQRYEKEIQ 408
 Db 341 R-YCSRNNGDCEKTRAKGLIRYKQKQCYLACNPYVDWINNKEQDQKQKRYDERIK 399
 Qy 409 SY-----LSNDKRVNNINSE--YKQFYEKIKETQYATNDFEMLNLEKRYC----- 454
 Db 400 KYENASGSGSNOKRQAGITTTNTDGYEKKFYDELINSEYFVDFKLEKLSNEEICTRYK 459
 Qy 455 --KGLPGEKDIITFNSA-----DCKGIFYRSEYQVCPDGCYKCDGITYTH-KSDND 504
 Db 460 DEEGTIDFKKNVNSDSTSGASTNVESEGTYRKRKYQPCPYCGVK---KVNNGSSNE 515
 Qy 505 RERVNN-----EDYKPPWGYKPTNITVLYSGNEQGITQKLENFCN-----SS 547
 Db 516 WEKNNKCKSKGKLYEPKPKDEGTITILKSGKHDIIEELKCFDEKNGDTINSGSG 575
 Qy 548 TNYKDKNN-----OKWECYKDENINRCKLEQNTFELNDNPK-----IT----- 586
 Db 576 TGGSGGNSGQELYLEKCK-YKGEDVYKVGHDDEDEYENKANNAGCLTILNOKKNKE 634
 Qy 587 -----SFHNEPELVYTLADTILKWNDKLTCTI--NNTTHC-IDECNRNC 629
 Db 635 EGGNTSEKEDLOKTFNPFYVVAHMLKDSIHKKKLQRCLONGNRILCGNNKKNDC 634
 Qy 630 LCFPRWYKQKEEENSTIKKLTETKKNQO-----QSYSNINNFEGYF---672
 Db 695 ECFRWRITQKKDEGKIVQHF-KTONIKRGSGSDNTAELLIPFDHYVLOYNLEEFELKGD 753
 Qy 673 -----FKVMDKLDDKAEKWELENIRKKNFESLNENRD-----YLENATELL 718
 Db 754 SEDASEEKSSENSIDAEBR-----EELKHLREIIESDNNQASVGGVYEOKNIMDKLL 807
 Qy 719 DHEKETATC-----KDNNTACETISHNATTPCVRKPGTQ-PTYNIKEIAQ 766
 Db 808 NYEKDEADLLELHEDEBEKEKGDGNECTIEGENFRYPNCSGSEGNKRYPLANKVAYQ 867
 Qy 767 YFKRSAYEARNRGLHLKLGKAHEGILYKRGRRKDKDNLCRIMHSNRNIGFSNGPCD 826
 Db 868 MHHRAKTQLASRGRSALRGDISLAQFNKRNSTLKGQICKTNEYNSDSRNSGPGPT 927
 Qy 827 GKGSGGICQIRFVVGEBWEKDPHMKRKHEDVIMPRRHICTSNLEHLQTDHPLNGNI 886
 Db 928 GK-DGHDGVMRKIGTWEWSNIEGKQTSYKNVFLPRRREHMCSTNLENDV-----GSV 980
 Qy 887 V-DDLVNNSFLGDVLSAKYEAANKILRMTEKKNKLGPEVNDP---KQITTCARIRS 942
 Db 981 TKNDKASHSLGBOVQAATDAEITIKRYKDONNT---QLTDPLOQKQOEMACRAVRIS 1036
 Qy 943 FADIGIILINGRIDWERNGDMVLOGLHETVFGNIHSLKG-KGNDKYNDP---ARKYLU 998
 Db 1037 FADLGIILINGRIDWMDKSDTDMETRLIVFANINKEKHGIDNKNRYTGDESKRAYYKL 1096
 Qy 999 RENWEMANAKYWEAKCKCIKYLKDKSGHQSOSTSGYSDHTPLDDYIIPOKLRMTEMA 1058
 Db 1097 RADWMAHNRHQAARAKC-----ATKGIIC--PMPYVDYIIPORLRMTEMA 1141
 Qy 1059 EWTCKQKQKYEYKLEKCEKCKDKNDNGOGCTKESGTGCKEFAENENDIIGLMEQWN 1118
 Db 1142 EWTCKQKQSOEYKIKKIKCADCKSKGDGK-CT-QGDYDCKCAACDKYKEETEKWNBQR 1199

QY 1119 IISDKTELHEAOMSVNSGIEASTAKNHID--RNVIETSELYOONGK-----SNK 1171
 DB 1200 KISDKTNLYIAQAKTTSTNG-----RTVLGDDDDPDYQOAWDELTPHKAISTARVLVRA 1255
 QY 1172 SGSDSAVIGTWTENTENGAYLHDGNDPDOSONEFODEK-----SDGONEXYAPR 1225
 DB 1256 AGSPTEIAAAPTTPY STAGYTHOEIGGCOEOTOFCEKKGAVSTSTTKNKEETFR 1315
 QY 1226 DRPODHDGACGCKSGSRPTFRVOJTKRKAENDTECKTYNDILKENDGKOVEDCHPKKN 1285
 DB 1316 OPEPEVATACDCINRSOTE-----EPKKKEENVESACKIYEKILLEGNGRTVECPKRES 1371
 QY 1286 SNGYPRMOC--GAINLYEDRVPCMPRRROKICVHFLANDNEIKLOSOVNLKEAFITSA 1344
 DB 1372 ---YPPMDCKNNDIHD--GACPPRRROKICLYIAHSEOTENIKTDNDKDAFITAA 1427
 QY 1345 EPEFSWYYSK--DGEENELDEKELKIPAPFRASMEFYGYRDLFTGDSKGE 1403
 DB 1428 EPELSMOYKSKNDSEAKILDR-----GLPSOFLRSMNTFGDYROICLANTDISKONDY 1483
 QY 1404 SKLEOIDSLEKNGDOKSPNGKTRQEMWTEHSHIWEAMLCALVIGAKKD--FTENT 1460
 DB 1484 AKAKDKIGKFEKSKSPSGISROEMWKTNGPEIKMGMLCALVYTDITDNKRKINDY 1543
 QY 1461 GYNNVFSDK--STTLEPAKRPOLFLMTEWYDYCYTROKYLKDOEKCS--NDOLKCD 1518
 DB 1544 SYDRVNOSONGNSPLEEFAKPOFLRMLEMGEFECAEROKKENIIKDNCNEINSTOOCN 1603
 QY 1519 T---ECKKCEDYVYKMK--KKEMIPDOKYKDERDKKRDROHIGVMTDYGTATDY 1574
 DB 1604 DAHRCNOACRAVOEYENKKEFSGOTNNFVLKANOVPDPEYKGYEND----- 1554
 QY 1575 LNRKFTASCGDKPSASVVOIRIOLLEKQAYYADRKHCCTKEIENDKYTNISDKCK 1634
 DB 1655 -----GVOP-----IGNEYLLOK--CDNNK--CSC----- 1676
 QY 1635 GLYKEANTGAIKMONKGNPNYNNLKELEVLPSRRLRICFALDONTDTPREYKDENG 1694
 DB 1677 -----MOGNVLSVSPK-- 1689
 QY 1695 RKRLMEVAATEGYNLGOYKKEKKEKIKTSDAHKYSYEVEPCSAMKSYFDRLITGI 1754
 DB 1690 -----PFGKY-----AHKYP----- 1699
 QY 1755 DNEDEKOTEENLKKIENKNGTSVGKSDSTGNGPSTARKEFWNENKECVNNAITGY 1814
 DB 1700 ----- 1699
 QY 1815 KRGRDONGNSARSDEDLKCKGSPSDDYPMKKNRDEBTAYOFLWFAWGEDEPCKH 1874
 DB 1700 ----- 1699
 QY 1875 EKELEKLVGACNDYTCGDNEDKRRKCTDACTQYKKEFISEWKPOYEQIKKYGKDIYS 1934
 DB 1700 -----EKC----- 1702
 QY 1935 EHPVAKDAEDAREYLDKOLKICENKSGDEYKCMKDVSTQRLTNGNSQNPASLDEPK 1994
 DB 1703 -----DC----- 1704
 QY 1995 EYEGKNCQYPRGPRPRRRETPSPRVLISKATASKEATAPPTKPKVENITTEMRA 2054
 DB 1705 -YOG--HVPSP--PPVPO-- 1723
 QY 2055 QTRRRAAQT RKRRTATTESDVTGVTMKAISNKPDSRGIGCPKPYGQ--PKWG 2112
 DB 1724 -----PPAPYTVDVCSIVKLEK--DTNNFSPACGLK--YGTAPSPSK 1764
 QY 2113 CI-----VGSKEENENGICMPPRRRKLCINNIOYLANET-----ENKRD 2151
 DB 1765 CIPSDTKSGAGATTKSGSGSISICIPRRRRRLYVGLQ--EWATALPOGEGAPSHSRA 1822
 QY 2152 NDKEARIKCAITFQPLMKIIE--NPAENE-----LONGT 2188

DB 1823 DDLRNFIOAALTEFFLDRYKKEKPPGDSGOAL.SOLTSTYSDEEDPPDKLLONK 1882
 QY 2189 IPDEFRIWYTYGDKXDFEFGTIDSNKKIITVNSVTTLNENKROKODKDELEK 2248
 DB 1883 IPPDFILMEFTLIGDRLIVHOGNTSDSGNTNGSNNNIIVLEASGNKEDMKIOKIQ 1942
 QY 2249 I-----FWENKRFIWEIGMIGLY-----HLTENEKIR 2280
 DB 1943 ILPKNGTFLVPSASQOTDKMNEHAESIIMKICALITYTEKNPDTARGDENKIEKD 2002
 QY 2281 DNY-----QYN--DMKL-----TPSLEEFVPR 2302
 DB 2003 EYVEKFFGTADKHGASTPTGTGYTOYDEKVLKEDTSGAKTPASDPLPLSDVLRP 2062
 QY 2303 OPLRFTMAEFCNRRKREOLLEIAGCKEY-----CN--GSN 2339
 DB 2063 PYRVLIEBMOGNCRRKRLKHAQIKHECAYEEBGGSSRRGIGITROYSGGECANEMLPKN 2122
 QY 2340 DGKTOE-----CAEACVYTONFIKKMTEYEROKERFKDKD--GKKRYDPTSTERTI 2390
 DB 2123 DGTVPDLERPCSAKPCSSYKWIIESKGEKEKAYEQOKDCVNGSNKHD----- 2174
 QY 2391 EKATCAHEYLMMKLELCGKDCSCMOKPSOLPKTTOOSOSDANDMESLDYVEEEN 2450
 DB 2175 -----NGFCETL-----TTSKADPLKTLG----- 2195
 QY 2451 KCEPELSKKGSMI-----HTKKITEP--KIPMVCAYEAYLSKEANNMITLE 2500
 DB 2196 --PCKNNVNGKITPDDKTFKHTKD--CDPCLKFSVNC----- 2230
 QY 2501 KFIPIESTEKESKNSWNTNPNCPDKPKPYAPDKY IGRNPNENRENERKDYEMKCYAN 2560
 DB 2231 -----KKECDNS-----KGTDCRN----- 2247
 QY 2561 SKFYOKKRRKCVPRRHHMLRLDEKIERLKDMSVLLKVMRTARNEGIDILKNFNS-- 2619
 DB 2248 S-----IDATDIENGVSIVLEKRSVADSLS-----GFNGD 2278
 QY 2620 --ENGAMNFICTMYKSFADLDIYRGTDMLRIGLYPEVETIKLYVEYIYKWRNNK 2677
 DB 2279 GLENAC-----RGAG-----IFGI----- 2293
 QY 2678 KGRNKYNDVOTFRSAMWANDRKDIKA-----MTCKAPDAKLFKGRMDGERLTLIO 2731
 DB 2294 -----RKDEMKCRNVCYVYCK--PENVNGEAKG-- 2321
 QY 2732 DKCGHDDPVDYIPIORFRMTEMSEYCYCKALMELEKFKSCDCKTSDRCNDYDEN 2791
 DB 2322 -----HIIOIRALVRYVEYF--FEDINIKIKH--KISHRIKN--G 2355
 QY 2792 KCEOCKRCOEYKRNFLKMSLFDIOSNKXYKELYEQPIYKISTYDHOVNFVOKLTFKS 2851
 DB 2356 EISPC-----INCYEKM--VDOKKEKKEKELTER-----FKD 2385
 QY 2852 ECYVESFSEYLLHETSKCLANTKENBNDGSSINRIYAAFEETPK-----SYKEA 2897
 DB 2386 Q-----YK--NDNSDDDVRSFLETLIPOITDANKKNVIRLSFGNS 2426
 QY 2898 CSCTLPKKNPLNCPDNDGCKELOTFFPGSKNDYDNNLDMNNAVILVNSSDNKGYL 2957
 DB 2427 CGGSASA-----NQNONG----- 2440
 QY 2958 IPPRRRLHCTRPITAYNYRRGRGDEILKLLTSAFSOGOLLGOKYKSEBELCEAMKYSY 3017
 DB 2441 ----- 2440
 QY 3018 ADVSDIITKGMADMTSLSEKIKIIFETSNBATERKRTWMENNRROIWHAALCGKIATFSK 3077
 DB 2441 -EYKDAI--DCMLKRLKDKI----- 2457
 QY 3078 VTLDEGMCQPLPROBETNQPLRWLLEMAKQACEKHHVSDSLKTRC-----PRSNEDNFEA 3132


```

Db 2458 -----GEC-----EKKHQT-DTECSDTPOPTLED----- 2483
QY 3133 SELLROPGCONDIRKYISILIKNTMENLNIYKOLKDOSSGNIDNKSEENVOSYIKS 3192
Db 2484 -ETL-----DDLETEBAKKNMFKICEV-LKTAQOEDE-GGCVAAENSEEPA-----A 2530
QY 3193 KDSQCALELNDINEIYTGKNNENNEFEKLVKLYPGLYFVEDHFKHVLGNKEEKO 3252
Db 2531 WDS-----GKETPQOTPVLK-----PEERA 2550
QY 3253 TVRPAKLYFTFPHVDSFOAPLEFSTRVAQYOPKN-----DLKSSISVYVSAVG 3303
Db 2551 VDEPP-----PPPOEKAPAP-----IPQOPPTPTQLDNPVHTALVSTLAMSVG 2599
QY 3304 L-----TALHPMKKKFKSSV-DILRLINIPQEGEYGMPTLSKNRYIPYRSGPKGYIYM 3358
Db 2600 IGAFTFYIYKKTSSVGNLFQIOLIPKSDYDIPFKISPNRYIPYTGKTRGKRYITL 2659
QY 3359 EGDY-----SGDEKYMMDLSSDITSESEYELNDIYVPSPKYKTLIEVLEPSKRD 3415
Db 2660 EGDSTGDSGYTDHY-----SDITSESEYEEMDINDIYVPSPKYKTLIEVLEPSGNN 2713
QY 3416 IFSD-----DTPS-----NDTPRTNFIIDENNELKHDVSOYL--- 3449
Db 2714 TTAAGNNTTASGNNTTASGNKPTSDTQNDIQNDGIPSSKITNEMQLDEFISQLOSE 2773
QY 3450 PTEPNNTKASADIPMTEPNTLYSDNPEKPPHISIHDRDLYTGKEISY----- 3499
Db 2774 PTEEP--NMIGYVNDNNHTPTSH--HNVKEKPFIMSHDRNLFSGEYEVYDMFNGSNMPI 2830
QY 3500 NTNMGNN-----TNNDIPMARNDSYRGIDILNDSLVYLNLLY--MAK 3541
Db 2831 NISDSTNSMDSLTLSNHSFYNDKNDLYSGIDLINALSGNHIDYDEMUK 2880

RESULT 4
ID AAM00384 standard; Protein: 2913 AA.
XX AAM00384;
AC AAM00384;
DT 21-FEB-1997 (first entry)
DE Plasmodium falciparum erythrocyte membrane protein.
XX Plasmodium falciparum; erythrocyte membrane protein; malaria;
KM detection; identification; treatment; prevention; parasite.
XX Plasmodium falciparum MC type.

Key
FT Domain location/Qualifiers
FT /label= Duffy binding ligand domain 1
FT /note= "Cysteine rich motif"
FT /note= "Cysteine rich motif"
FT /label= Duffy binding ligand domain 2
FT /note= "Cysteine rich motif"
FT /label= Duffy binding ligand 3
FT /label= Duffy binding ligand 4
FT /note= "Cysteine rich motif"
FT /note= "Cysteine rich motif"
FT /note= "putative transmembrane domain"
XX W09633736-A1.
XX 31-OCT-1996.
XX 26-APR-1996; 96WO-US05798.

```

```

XX 27-APR-1995; 95US-0430908.
PR (AFVY-) AFFYMAX TECHNOLOGIES NV.
XX Baruch DI, Howard RJ, Pasloske BL;
XX WPI: 1996-497376/49.
XX N-PSDB; AAT41852.
XX New plasmodium falciparum erythrocyte membrane proteins - used to
XX develop products for the diagnosis, treatment or prevention of
XX malaria parasite infections
XX Claim 1; Figure 12; 149pp; English.
XX A polypeptide comprising a plasmodium falciparum (Pf) erythrocyte
XX membrane protein 1 (PfEMP1) or active fragments or analogues of that
XX protein can be used in the treatment or prevention of symptoms of a
XX malaria parasite infection. The polypeptides can inhibit, block or
XX reverse the sequestration of erythrocytes in patients suffering from
XX malaria. Nucleic acids derived from the PfEMP1 gene can be used as
XX probes and primers to identify a plasmodium falciparum parasite, the
XX primers used to generate characteristic amplification patterns from
XX different P. falciparum strains. Antibodies specifically
XX immunoreactive with the PfEMP1 polypeptide or its fragments may be
XX used in diagnosis of malaria infection. This is the PfEMP1 protein
XX used in the MC type of plasmodium falciparum. An alternative, truncated
XX PfEMP1 protein is given in AAM00385.
XX Sequence: 2913 AA;

Query Match 16.0%; Score 3108.5; DR 17; Length 2913;
Best Local Similarity 26.4%; Pred. NO. 1.9e-177;
Matches 995; Conservative 446; Mismatches 963; Indels 1367; Gaps 150;

QY 33 KSARNVETRYAKNI-----RHPSRYAKEHYDSLKGLDTKAERFGSPSTPVNKHNYYPY 86
Db 13 KDAKHALDRICEEYKKEVENDAEKRYK--ALKGNLQEAKEGIGELASS-----PN 60
QY 87 PCNLDHKEHTNLRYDDVNLRHPCGREGONRFEDESEBEG--NKIRNRYKRNDAL-ACAP 144
Db 61 PCKL-YEDYIYNNRLK--RKRRPCANRQTVRPSDEYGGQCTFNRIKSDENNDSIGACAPY 117
QY 145 RRRHMKCNLEALNDINTONIHDLGNVLYTAKYEGESIVNHN--PHKGT-----SDCTA 198
Db 118 RRLHICYNLEKMKRTSTTK--HGILLDYCAAKTEGDSIKTHYKHELTNPDTKSQLCY 176
QY 199 LARSPADIGDIVRGIDMF-----KPNVHDKVETGLREVRFKIHGMEDEVKNDYNPDGS 252
Db 177 LARSPADIGDIVRGIDMF-----KPNVHDKVETGLREVRFKIHGMEDEVKNDYNPDGS 252
QY 253 GNYKKLREAMNVRNKNVWEATITCDASYKSGYFPMQSSNTPPLSNP--KCHKROGK----- 306
Db 237 ENTQOLREDMWTANRSYVMAITCHAGESDKYFEKTCSCSEWTDKRCRDEGSKETNE 296
QY 307 VPTNLDYVQYLRWFDEMGEEFCRKNRIKLYKDSGRNDK--ERYLCSHNGHCTTTIY 364
Db 297 VPTYEDYVQYLRWFDEMGEEFCRKNRIKLYKDSGRNDK--ERYLCSHNGHCTTTIY 364
QY 365 KKGILHDKNCTDCSTCKYFEVWLGNOQAEAFKKOKYKKEIQ-----SYLSNDKNFVN 420
Db 356 GAETVEYKGDCHGCSVACDFRVKIDNOREKPEFKOKYKKEIEINKGHTITGNGKINN 415
QY 421 INSEYKQFYEKLKETQYATNDTFLNLTNGCKYK--GGLPGEKDTLTFNSADKGIYR 478
Db 416 L-----YVHFHFKIILK--YTPYVDSLOKLTNDEALCKRPVNGNEKASTVDFNEVNTFSH 471
QY 479 SEYQVCPDCGVKCD-----GIKTYHKS--DNDREVRNNDYKPPWGVKPTNIVLYSGNED 533
Db 472 TTYCEACPMCGAQKREKNGGMAKESKAKKERIRFKE-----NSDIDIKILTFEKGK 524
QY 534 GDITQKLEFNCNSTYKKNKNNOKMECIYKDEININCKLEONTIINN-----N 582

```

Db 525 SKTELEKTFCKADGOKIK---NDIKCHDDNG-----TDDOTDSDNDVLDGMDGLTE 576
QY 583 PRIISFHEFEELWYVLLYLDJTKWMDKLCTGIN-NTTTCIDECNBNCLCFPRWVQKEE 641
Db 577 DKIMSYNAFEMWVHMDLDSIKWDEHGRICINKDKGKICINGCNKKICFCQKWVQKKT 636
QY 642 EWNSTIKLTKKKNKNOQSYYSNNLFEQYFFKVMKLDKDEAKKMELENTIKRKNES 701
Db 637 EWGKIKDRKROKDIJPKOM----- 655
QY 702 NLENNRDYENAI--ELLDLKETATICKDNNTNEACETSHNATNPCVPRGQTPIK 759
Db 656 ---THDFFQTLMDKDLLEIIOITY---GDANE----- 683
QY 760 NIKETIAQYFKRSAYEBARNRGLHKLKGAHEGIYKRGGRKDKDNLCRIMIKHSNRNLG 819
Db 684 -IKRI-----EALBGA---GVGGIDFALAGIYTGFAVE- KOTITDKLQHEBK 729
QY 820 FSNGPCDGGTGDIGIOTREYVGTMEVDEPHARKHEDYIMPPRRRHICTSNLEHLQTD 879
Db 730 ---EADKCLKTHTDYCPPO-----ED 748
QY 880 HPLNGNIVDVLNNSFLGVDLISAKYEANKIIRMYEKNNLKGPRKVEDPKHOTICRAI 939
Db 749 RSV-----ANSESATV-----PSPPADPK----- 767
QY 940 RVSFADIGIINGRDLMEHNGDMVKLOCHLETVFGNIHSLKGNKNDKAPKYLKLR 999
Db 768 ---ATEVDANASDGEDDEE----- 787
QY 1000 EKWMEANRAKWEAMKCDIKYLKDSGHOSSTOSSYCSIDHPLDIDYPOKLRMTAE 1059
Db 788 --- 787
QY 1060 WYCKVOKEKYDLKECKCKNDKNGOGCTK---ESGCTCTKTEACNENYNDIIGLMEQ 1116
Db 788 ---EEDEDEGEBAEVEQEKDESAATEVAAPSPPTTO----- 824
QY 1117 WNIISDKYELHEQOAMSVNSGIEASTAKNHIDRNVLEPSELVQOONGKSNKSGTSD 1176
Db 825 ---DGVPAQEDDVKCSI-----VDKAL-----KGLD 851
QY 1177 ESANVT---GTNTYENVAGALYHDIGNFDOSQNEFCOEKSDGKNEXAFADKQDDH 1232
Db 852 DACTLKYGKGDYTTTEST-----TKP----- 872
QY 1233 GACGCKSGSKPTFVOJKTKKKAERKDECKTYNDILKENDGKQVEDCHPKKNSNGYDPM 1292
Db 873 GAAGTSG-----KDT----- 883
QY 1293 OCGNINLVEDPBYCMPDRROKLCVHFL-----ANDN----- 1323
Db 884 ---GSI-----CVPRRRKLIVYGLHDMAGGETTEAKSOETSOGOKTPSGNESSPSRK 933
QY 1324 ---EIKKLOSQVNLKELAPLKSAAETFPSTWYYSK-----D 1357
Db 934 LPOGPPTETTKETPESSLJHAFFVSPRLRFLPWHKFKKQMAQOAGAGATGOOTIITCLD 993
QY 1358 GEGNEL--DKELKEGKIPAPLBSMEYTPGDYRDLFG--DI---SGHGEGSLKQJDS 1412
Db 994 GGGEEPLDLKLTGHIPLPDLKROMFTYLDYRDLVGNNDIYVHTSGNEDMOIMEAIQK 1053
QY 1413 ---LFKNODOKSPNGKT-----ROEWMTSHSEIWEAMLCALV----- 1447
Db 1054 KIRQIILPTSGSSPSPPRYQTOHSHVENPKRTWMNENGGKIMEGMVALYNTDTPSGTAP 1113
QY 1448 ---KIGAKK--DFTEN-----YGINNVKFS-----KSTILEEKKRPO 1482
Db 1114 TOIOEVRTIKMDENSNPKPIPOYKYDOVKLDPTSDAKTTGSPISPSEKKTPLDFTSRBP 1173
QY 1483 FLRLMEWYDDYCYTROKYLKDOVECK-----SNDQLKC-----DPECNKK-----C 1525

Db 1174 YFRYLEENGETPCKEKRRKLEKICECRDRTGHEHSGSDGIDCTRTDADRDKFPYDNL 1233
QY 1526 EDYVYKMKKKEEM--IPDQYKXDERDKKRFDRONIGVWTDYTGNAADYLNRRFTASC 1583
Db 1234 ROCHIOCRKRYRRWIDIKFDEYHKOEK-----KYOG-----EY----- 1265
QY 1584 GKRPASAVVQNRNIOLEQAYYADAKHGGCTKFIENDKYNISCKOKGLYKAEANTG 1643
Db 1266 -DK-----LTKDKSSGGDNMC-----CKDIEKHKSAA 1291
QY 1644 AIKMQKGNPNYNNLKELEDEVLPFSRLRICEHALDGVNTPDEVENGRLRLMEVAA 1703
Db 1292 VF-----LKEK-----HCNGOTSEKKGQEDOL----- 1316
QY 1704 TEGYNIGOYKKEKEKEKI-KTSDAKHYSYEVPPCSAMKYSFYDLNDIILIGDNLDEKO 1762
Db 1317 ---NKLDPRDIPOFSPSTYKACB-----YGV----- 1342
QY 1763 KTEENLKILFNKNKGSVSGSDS--TGNPSTARKFNNENKECVWMMIC-----G 1813
Db 1343 ---NCGNKRGGGNGCTTNNB-----PENKENDGGAASITISILINDG 1383
QY 1814 YKGRDDGNSGNSAREDEDLKKGVSDDDYPM--GNRRDEGTAYOFLRWEAWEDEDC 1871
Db 1384 STINGATNGTGT---TDETLKEC---SDKTAFFGLKROEYTCOK-----KTVNGC 1429
QY 1872 KHKKELEKLVGANDYTGDNEDKRRKCTDACTQYKFEISEWKPOYERQIKYGENKOR 1931
Db 1430 ---NLJNRVND--TYFDKO-----IYNEFEFORLRYF---VHDYNLKHK 1467
QY 1932 IYSEHVAADAEDAREYLDQKLKICEKNSGCEYCMKDVSTQRLTDGNSOMPASLDD 1991
Db 1468 I---DPCIKKEKODKT---EHKCI-----NGCNI----- 1490
QY 1992 EPKEVEKCNQVPRGRPRRRETSPRVLISKATASKAEAKTAPTKQPKVENULTTE 2051
Db 1491 ---KEEC-----VR----- 1496
QY 2052 MRAQTPRRAAOQTRRTSTATTESDVTGMVKAISNKPDSHGIEGCPNPKTYGOYPRW 2111
Db 1497 --- 1498
QY 2112 GCIYVSKENENICMPRRKILCINNIOYLANETENKNDNDIKELFICAIETQFLWL 2171
Db 1499 LELKG---NEWG---NKR---KHNINSNDKET----- 1523
QY 2172 KYIIEPAENELQNTIPDEFKRYMYTYGYDKDM--PFGTDSNDKCIITVNSVTIL 2230
Db 1524 ---LAINVKSIFYVDGLFDIDYKKAQVY----- 1549
QY 2231 NENNRKODKKDELEKIFWEKNNKFIWEGMAYGLTYHLTDE--NEKERIRDNYOYNDMT 2289
Db 1550 ---EDEKERRKIM-----GCTGH--DECSKKEBENK--NFTI 1580
QY 2290 KLPISLEEFYKRPQPLRWFTENAEFCNRKQEDLLKLEAGCKYEBGNSNDGTCQACA 2349
Db 1581 NLISLQDKITSCQ-----NKH-----NPNKGT-----A 1604
QY 2350 CVTYQNFKKMTEYERQREKFKKDKGKKYKDYPTENDIEKATCAHEVLMNKLKELOG 2409
Db 1605 C----- 1605
QY 2410 NKDCSCMQRPSSQLPKTQOSOSDANDMPESIDYVPEEFNKECEPDELKSKGMIHTKKI 2469
Db 1606 ---DPEPSPTEBETDPLDDPDPDPLDDDOHTGQPKFCP----- 1640
QY 2470 TEPKIIMNCEYKAYILSKAEENNDITLKEK-----FPIESTK--EK 2511
Db 1641 -PPPPMTCVEKIAKELRYEAEGKINNELKGNCKDENGCNNYKKNAGAVIGESCKPFO 1699
QY 2512 ESKNSWTNNNPPCDPKFVAPDKYIGRRNCPENNEERFPVDEYMKCYKNSKFOE--KKRY 2570
Db 1700 TYENSYNIN-----NKCKDNQNERFKIGKW-----NFKIIGITRKDI 1738

```

OY 2571 CVPREHMCIRNDEIKERLKNOSNYLLKAVBRTAREGIDITKNFNSGNCAMPICD 2630
DB 1739 CIRPREHMCIRNDEIKERLKNOSNYLLKAVBRTAREGIDITKNFNSGNCAMPICD 1798
OY 2631 TMYKSFADIGDVRGDMIRIGLYLPVEIKLYKVEFYIGKWRN-KN--KGRNKYNDV 2686
DB 1799 AMKYSFADIGDVRGDMIRIGLYLPVEIKLYKVEFYIGKWRN-KN--KGRNKYNDV 1858
OY 2687 QTFRSMWANDAKDITWAKMTCAPDAPLFRGRND-----GFEITLLOKCGHKDPP 2741
DB 1859 ---RSDMDANDAKHIMWAMTCAPDAPLFRGRND-----GFEITLLOKCGHKDPP 1915
OY 2742 VDDYIPORFWRMTESEYKCALMELEKFKSCDHC-KTSDRCNDYDENKCEQCKTRC 2800
DB 1916 DYDITPOFPKRWQWSESEFCILINEMEQFEKTCGCKKNSTICGDDNNGTCENCKNQC 1975
OY 2801 QEKNFVLKWSLFDIOSNKYKELYOPIY---TKISTYHVQNFVKLTFKSEC-SV 2855
DB 1976 EYKKLILHMKLGFD---KYKEIYNE-IYNNKDSKINSNEYKFKLEKL---KDKCKEL 2027
OY 2856 ESEFYIHTSKCLMKNKRENDGSSNIRYAFEBETPKYKCEKSCGLTSLKNPLDNCPIQD 2915
DB 2028 NSSDCIDEATICTRYKFS-NSEKNHNNYAKENPKKEYEKACKDAP--DPLDNCPRDS 2084
OY 2916 --NKDGCKELQTFECSKNDYNNLDNNNAVLYLNSSDNKGVLIPPRRHLCRPIYAY 2973
DB 2085 ATYERACHTLPTKICESKTFENNDDSDWTSFVOTSPRDNIGVLVPPRRROJCLANKITT- 2143
OY 2974 NYRKGRD-EILKKLLTSAFOSOGOLLGOKYKSEELCEAMKYSTADYSDITKGTDMADT 3032
DB 2144 KLRSEKIDIDFKAELMTSYNEGKLCILCELYKKRDVYTLQAMKYSYDYDYGKGLDILST 2203
OY 3033 SISEKIKKIF-----EISNEATEENKRWENNRROIWAMLCGYKIATSKYTLLEGMCQ 3086
DB 2204 APLDKLTKLNLVLAGDGTNEJKEKDRGKMTENKTRVWAMLCGYKIATSKYTLLEGMCQ 2261
OY 3087 LPRDEETNOFLMILWAMKQACKERKHSVSLTKQPSRNEDEFASELLRPGQ--N 3143
DB 2262 LPRDEETNOFLMILWAMKQACKERKHSVSLTKQPSRNEDEFASELLRPGQ--N 2318
OY 3144 DIRKYSILNLTAKNMENINITYKOLKODSSGNIDNKPE-ENVOYSIKRSOCALEN 3202
DB 2319 QYRDYITRKIQ-----EYRLNLYQYNTNFKKAEVTKAPYENDK---CNDKCN 2365
OY 3203 DINEIYTGKNNENNNEFKVLLKLYPGVLEDETHKNHVLGNIKEEQYRPRALYFE 3262
DB 2366 CLSKYT-----DIEKKKMMYVSF---DDNOLKNCICROIKPRP---PKVKPE 2410
OY 3263 TPRHVSFYQAPLFSTHRAQYDP-----KNDILKSSISVIVSALGIALHFMKKFKFS 3316
DB 2411 EHEPTSEQPTPLPPKPDLPPEAPENRDLLEKTIPEFGIALMAGIAFLFLKKTKFS 2470
OY 3317 SV-DLRIINITQOEGMFTLESKNYIYRSGPYGKRYIIMEGT---SDEDEKYMMD 3372
DB 2471 SVGNLFQIILHPRSDIDPTKLSPNRYIPTYSKGRKRYIILEGSGTDSGTDHY--- 2527
OY 3373 LSSSDIT--SESSEYELDINDIYVPSPKYTLIEVLEPESKRDIPSD--DTPS----- 3423
DB 2528 ---SDITSSSESEYELDINDIYVPSPKYTLIEVLEPESKRDIPSD--DTPS----- 2584
OY 3424 NDPTPTNFFIDENNELHDFVSYTL---PTEPNNTKASADIPNTEPNLTLYSDNPEE 3479
DB 2585 QNDGIPSSKLTIDENWNTLQDEFISMLQNEPNTPE--NNMIGYVNDNTPPT--SRANVEE 2641
OY 3480 KPRITSIHRODLYTGKEISYINIMSTNTNDIPMANRDSYIGDILINSL 3530
DB 2642 KPRITSIHRODLYTGKEISYINIMSTNTNDIPMANRDSYIGDILINSL 2689

```

RESULT 5
AAW22482
ID AAW22482 standard; Protein; 2710 AA.

```

XX AC AAW22482;
XX DT 07-OCT-1997 (first entry)
XX DE Plasmodium Proj3.
XX DB DL gene family; SAMP, sialic acid binding protein; vaccine; therapy;
XX KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
XX DBAB; mezoizolite; malaria; var-1; var-2; var-3; var-7; immune response;
XX KW Plasmodium.
XX OS Plasmodium falciparum.
XX PN WO640766-A2.
XX PD 19-DEC-1996.
XX PF 07-JUN-1996; 96WO-US09508.
XX PR 07-JUN-1995; 95US-0487826.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Chitins C, Miller LH, Peterson DS, Sim KL, Su X;
XX PI Wellens TE;
XX DR WPI: 1997-052231/05.
XX DR N-PSDB: AAT72897.
XX PS New malaria vaccines - contains cysteine-rich DBL family protein
XX binding domains homologous domains of the Duffy and sialic acid
XX binding proteins
XX PS Disclosure; Page 50-56; 96pp; English.
XX CC This sequence represents Proj3 of Plasmodium. Proj3 belongs to
XX CC the Duffy binding like (DBL) family of genes which have homology to the
XX CC Duffy antigen binding protein (DABP) and sialic acid binding protein
XX CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
XX CC var family of genes modulate cytoadherence and antigenic variation of
XX CC Plasmodium infected erythrocytes. SAMP and the Duffy antigen binding
XX CC protein (DABP) are soluble proteins that appear in the culture
XX CC supernatant after infected erythrocytes release merozoites. DABP and
XX CC SAMP mediate the binding of merozoites and schizonts to the erythrocyte
XX CC surface. These proteins are necessary for erythrocyte invasion by the
XX CC parasite. This sequence can be used in the compositions of the
XX CC invention. The compositions are for the treatment and prevention of
XX CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
XX CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, SAMP.
XX CC Family of genes having homology with conserved regions of DABP and SAMP.
XX CC The compositions are used for the treatment and prevention of malaria.
XX CC They are also used in the preparation of vaccines for inducing a
XX CC protective immune response in a mammal to Plasmodium merozoites
XX CC (especially Plasmodium falciparum or Plasmodium vivax).
XX CC
XX SQ Sequence 2710 AA;
XX
XX Query Match 15.3%; Score 2972.5; DB 18; Length 2710;
XX Best Local Similarity 26.5%; Pred. No. 2.6e-169;
XX Matches 865; Conservative 442; Mismatches 873; Indels 1083; Gaps 120;
OY 9 IIRKGNAASSLEGGAKSPITIKESHKASARNVLERAKNIRHPSK-YAKHHVLSLKGDLTKA 67
DB 6 LAKMPKRAAGDDIED-----ESAKHMPDRIGKDYDVYKKEKRGKGLGRISA 58
OY 68 EFRGSPSTPVNKHNNYYPYPCNLHKEHTNLRYDDVNLRHPCGHEGRKRFDEDESEC-G 126
DB 59 KFEKNESDPQPED-----PCDLHRYHTV---TTVINACADRSVRFDEDEGQCTH 110
OY 127 NKIRYKRRKNDIACAPRRRRMCDKNLEALNDINTQNIHDLGNVLTARYEGESIYNN 186
DB 111 NKIKSQGGDNKACAPRRRLHVCQNLQETEPITITHTNLVLDVCAAKFEGQSTIOD 170

```

QY 187 HP-HKGT-----SDACTALARSFADIDIVRGIDMFKPNVND-----KVTGLREVEREKI 235
D 171 YPKQOAYGDSPOICTMILARSFADIDIVGRDLYLGNFOEIKOQROOLENNILKTIKGI 230
QY 236 HD---GMEDEVDKNDYNDGSGNYKILREAMNVNRKVTWALITCDASYKSGFYQMSNT 292
D 231 YKINGLAEARYGND--PE---EFKLEDMWTANRETVMATYCN--WGNTYF-HATCNR 282
QY 293 PLESPNACGHKOGKVPNTLDVPOYLWMPDEMGEEPCRKRIKILKAKYKDCR---NDKE 348
D 283 GERKYGCRONDQVPTFYFYVPOYLWMPDEMGEEPCRKRIKILKAKYKDCR---NDKE 342
QY 349 RLYCSHNGDCTTTIMKGLIHLNDKCTDSTCKYKVEWLNQOZAFKOKERYEIQ 408
D 343 R-YCSRNGDYCEKTRAIKGLRYKOCISCLYACNPYVDMNOKEOFDKOKKKYDEIK 401
QY 409 SY-----LSNDNKFVNNINSE--YKQFYEKLETOYATNDTEFLNLEKYYC----- 454
D 402 KYENGASGSGRORAGGCTTTNTDYEKFFYDELNKEKRYVDEKFEKLSNEEICTYKVK 461
QY 455 --KGLGPERKDITFTNSA-----DDKGIYRSEYCOVPCDGVKCDGJIKYTH-KSDND 504
D 462 DEEGTIDFKNVNDSISGASGTWESQGTFRKCYQPCPYGCVK---KVNNGSSNE 517
QY 505 RERYNN-----EDYKPPWGVKPTNITVLYSGNEQDITOKLENECN-----SS 547
D 518 WEEKNNGCKCKSGKLYEKPPEKEGTTITILKSGKHDDIEELKFCDEKNGDTINSGSG 577
QY 548 TNYKDKNN-----OKMECYKNDENINRCKLEONTEIINDNPK-----II----- 586
D 578 TGGSGGSGRORAGGCTTTNTDYEKFFYDELNKEKRYVDEKFEKLSNEEICTYKVK 636
QY 587 -----SFHNFELVWLYTLNDITKMNKLTCTI--NNTTTC-IDEENRC 629
D 637 EGGNTSEKPEDELOKTNPFYFYVVAHMLDSIMKKKILORCLONGRIRICGNKKKNDCC 696
QY 630 LCPDRWYKOKEBEWNSTKILFTKKNKIQ-----OSYNNINNEFGYV--- 672
D 697 ECFKRWITOKKDEWCKIVOHF-KTONIKRGSGSDNTAELIPFDHVLVOYLADEEFLKGD 755
QY 673 -----FKYVDKLDKDEAKKELMENIKRKKNESNLENNRD-----YLEVAITLL 718
D 756 SEDASEKSENSIDAEEA-----EELKHLREIIESDNNDEASVGGVTEOKKIMIKL 809
QY 719 DULKETATIC-----KDNNTNEACETSHNATTNCPVKRGGTO--PTKNIKEIAQ 766
D 810 NYEKDEADLCLEIHEDEEKEKGDNECIEBENFRYNCSSGSKRYPLANKVAYQ 869
QY 767 YFKRSATVEARNGLHLKGAHEGILYKRGRRKDFKDNLCRIMIKHSNRLGFSNGPCD 826
D 870 MHHAKTQILASRAGRSALRDISLAOFKNGRNGSTLKGOLCIENENSNRSGNSGSPCT 929
QY 827 KGTGSDGDIOTRFVYGTMEVDEPBEHMRKHEDVIMPRRHRTSNTLHLOTDHPLNGNI 886
D 930 GK-DGSDGGRVARMRTGTSMSINEGKOTSYKNVFLPRPREHCTISNLENDV-----GSV 982
QY 887 V-DOLVNSFGLDYLISAKYEANKIIMYKKNLKGPKVETDP---KHOTTICRAIYS 942
D 983 TKNDKASHSLDGYOLAFAKDAEIIIRKYDQNNI---QUTDPTQKDEAMRAVAYS 1038
QY 943 FADIGDIIIRGRDLMERNGDMVKLOGLHETVYGNHKSGLK-KGDKXNDND---APKTLK 998
D 1039 FADIGDIIIRGRDWMDEKSSDMDTRELITVEKNIKERHNDGKNDPKYTGESSEKPAAYKL 1098
QY 999 RENMEANRAKVEAMCDIKYLKDSGHSTOSSYSGSDHPPLDLYIPOKLEMMTEMA 1058
D 1099 RADWMEANRQVWRAKMC-----ATKIIIC---PGMPVDDYIIPORLMMTEMA 1143
QY 1059 EWTCKYOKKEVDKLEKCKECKDKDNGOGCTKESGTCTCTEACNEVNDIILGKMDWN 1118
D 1144 EWTCKASOEYDKLKICADCMKSGDK-CT--QGDVDCGKCAKCDKRYKEIETKMNQMR 1201

QY 1119 IISDKYELHDOQMSVNSNGIEASTAKNHID---RNVIETSELYOQNGK---SNK 1171
D 1202 KISDKYMLLYOAKTISTNG---RTVLGDDDDPYOQWDFLPIHAKASIAARVLKRA 1257
QY 1172 SGTDESAVIGTNTYENVCAYILHDPTGNFDDCOQNFCEBK-----SOKNDEKTAFR 1225
D 1258 AGSPTETAAAAPITPYSTAAGYIHQEIYGCGOQOTOFCEKHKHATSTSTTKEREYTFK 1317
QY 1226 DKPDHDCAGCCKSGSKRPTVQOITKTKKAEBEKDECTVNDILKEKNGKQOVEDCKHKN 1285
D 1318 OPRPEYVTAOCDINRSQTE-----EPKKEEVEBACATVEKILLEGKNGRTTVCNCNRES 1373
QY 1286 SNGTPDMQC--GNINLVEDPRVYCPRRROKLCVHFELANDNEIKILOQOVNLEKAFKSA 1344
D 1374 ---YPMDCCKNNIIDSIDH--CACMPRRROKLCVYIAHESOTENIKITDNLKDAEIKTAA 1429
QY 1345 ETEFSWYKYSK--DGEENELDKELBEKIPAPIRSFYFPGVDRDLDFGDISKNGEG 1403
D 1430 ETEFSWYKYSKNDSEKILDR---GLIPSQFIRSMKYTPGDRDCLNTDISKONDV 1485
QY 1404 SKLEQOISLFEKNGDOKSPNGKTRQEWMTESHEIMFAMICALYKIGAKKDD---FTENY 1460
D 1486 AKAKDOKIGKFFSKSGSKSPSGLSRQEMWKTNGPEIMKGMICALKYTPDNDKRIKINDY 1545
QY 1461 GYNNVAFSDK-STTLEBPARKPOFLRWLTWYDYCYTRQOYIKDVOEKCS--NDOLKCD 1518
D 1546 SYDKVNOSONGNPSLEEFPAKPOFLRWMIWEGEEFCAEROKENIIRDACNEINSTOQN 1605
QY 1519 T---ECKKCEDYKVKMK--KKEMIPQDKYKDRDKKRRDRROIGIYMWYDITGNATDY 1574
D 1606 DAKIRCHQACAOEYENKKEEFGOTNNEVLNANOPODPEKGYEYKD----- 1656
QY 1575 LNRKFTASGCDKPSASVONRQILEKQAYYADAKHCGCTKFIENDKYTNISSKDKK 1634
D 1657 -----GVQF-----IGNEVYLQK---CDNNK-CSC----- 1678
QY 1635 GLVKEANTGAIKWKONKPNNTNINLKEITEDVLPFSRLRICFHALDGNTYDPEYKDNEL 1694
D 1679 -----MDGNVLSVSPERK--- 1691
QY 1695 RKRLEVAATEGYLAGOYKKEKKEKIKITSDAKYSYEVPPCSAMKYSYDLADIIIGI 1754
D 1692 -----PRGKT-----AKYF----- 1701
QY 1755 DNEDEKQTEENLKKIFNKGTSVSGKSDSTGNGPSTARKFPMNENKDECVNNAICGY 1814
D 1702 ----- 1701
QY 1815 KRGRDNGNSGARSDEDLKCKGVSDDDYPMGKNRDEGTAVOFLRWFAWEGEDFCKHK 1874
D 1702 ----- 1701
QY 1875 EKELEKLVGACNDYTCGDNEDKRCCTDAGCYKKFISEMKPOYEQILKRYGENKKIYS 1934
D 1702 -----EKC----- 1704
QY 1935 EHPVAKDAEDAREYLDOLKLTICENKSGDEYKCMKDVSTORLTGNSQMPASLDEPK 1994
D 1705 -----DC----- 1706
QY 1995 EVEBKCMQVPRGPRPRRRETPSRVLSIKATASKEAKTAPPTKOPKKEYNLTTEMA 2054
D 1707 -YQCK--HVPSTIP-----PPPPVQPO----- 1725
QY 2055 QTRTRRAAQOTRKRRTSATTTESDVGTMYKAILSNKDSRSGTICGPKRYGOV--PKWG 2112
D 1726 -----PPAPYTYDVYGVSVITLFR---DTNNFSDAGULK-YGTAAPSWM 1766
QY 2113 CI-----VGSKENENGICMPRRRRIKINNIOYLVNLET-----ENKRD 2151
D 1767 CIPSDTSGAGATGKSGSDGSGISICIPRRRLVYGLQ--EWATALPQGGAPSHSRA 1824
QY 2152 NDIREAFIKCAAITOTLMLKTIIE-NPAENF-----LONGT 2188

```

1825 DLRNFAISAAIEFFLWDRYKEEKKFQDGSQAALSLQTLSTYSDDEEDPDKILQNK 1884
1889 IDDEFKRIYTYGYDKDMFEGDISNDKIIITVNSVTTILNENKRRKQDKDEELK 2248
1885 IPPDFLRLMEYTLGVDRLVHGNTSDSGNTNGSNNNIVLEASGNKDKIOKIO 1944
2249 I-----FWEKKKFIWBMGYGLTY-----HLDENKEKTR 2280
1945 ILPKNGTLPVPSAQTDPDKMWNHESIMKGMICALYTERKNPDTSAEGENKIEKD 2004
2281 DNY-----QYN-DMTKL-----TPSLEEVKRP 2302
2005 EYLEKEFGSTADKHGASPTGTQYQYKEVKLEDTSGAKTPASASDTPILSDVLRP 2064
2303 QFLRMTEAEFCNKRKQKLEAGCKEY-----CN---GSN 2339
2065 PYFRLYEONGNCKRKRKLAQIKHECKVEENGGSGRGITROYSGDEACNEMLPFN 2124
2340 DKRTQE-----CAEACVYQNFIKKKKTEYEROREKFKDKD----- 2376
2125 DDTVPDLKPSCAKPSYRKWIESKFEKOEKAYEQKDCVNGSKHNGFCETLT 2184
2377 -GKKYKYD-----PSTERDIEKATCAHEYLNMKLELCSNKCSCQMKPSSQLPTTQOS 2430
2185 TSSKAKDLFKTLGPKCKNNVEBKTIFFDDKTFK-----HTKDCDLKTSVCKR----- 2234
2431 OSSDANDMPESLDYVPEEFNKECPBLSKKSGMHTKTKTEPKIPMNCVEKAAVYLSKRA 2490
2235 -----DECD-----NSKGDCRNR-----NSID-----ATDI 2256
2491 ENNDITLKEFPIESTREKSKNSMTNNPCDPRKPYATIGRRPCEN----- 2543
2257 ENGCVSTV-----LEMRVASDSKSGFNCD-----GLENACGAGIFECI 2295
2544 REENRFKYDWEKCYKNSKFEYOEKRVCPPRRHHCLNLDIEKIERLKDGNYLKMYR 2603
2296 RMD-----EMKC-RNVCGY-----VCKP----- 2313
2604 RTARNEGIDITIKNFENSENCAMPICDTWKYSFADLGDIVRGTDMLRIGVLPVEIKLY 2663
2314 -----ENVNGE-----AKGHITQIRALVK-----RMV 2336
2664 RVEFYTGKRRKNGKRNKYND-----VOTFRSAMWDMNRKDIWAKMTCAPEADAKLFRG 2719
2337 EYFFEDYKIKIKIRKINGISPCIKNCYKWDQKRKE-WKETTEFKD----- 2387
2720 RMDGERITLLQDKCGHKDDPVVDYIPQRFKMTWSEYCYKALMELEKFKSCDHCK 2779
2388 -----QYKNDNSDDDNVRSFLETLIPQITDAN---AKNKVITLSKNGSCG-CS 2432
2780 TSDRKNYIDENKCKOCTRCOEYKNEFVLAKKSLFDIOSNKYKELYEODIYKISTYDHV 2839
2433 AS-----ANQONKN-----GEYKDAI-----DCMLKRLKD-----KICECEK 2465
2840 ONFYOKLTFKSECVSESESEYLHETSKLANKFNENDSSNIRIYAFEEPKSYKE--- 2896
2466 HH-----QTSOTECSDTQPOPLED---ETLDDDIETBAKKNMMPKICENILKTAQODE 2518
2897 -ACSTLPSKPNLDCPTDQNKD 2918
2519 GGC---VPAENSEEPAADTSGE 2538

```

```

XX DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KW DABP; Sialic Acid Binding Protein; SAMP; malaria; vaccine; immunisation;
KW protozoa; Proj3.
OS Plasmodium falciparum.
XX US5939827-A.
XX 30-NOV-1999.
XX 07-JUN-1995; 95US-0487826.
XX 10-SEP-1993; 93US-0119677.
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX Sim KL, Chitnis C, Peterson DS, Su X, Welles TE, Miller LH;
XX WPI; 2000-194198/17.
XX N-PSDB; AA298286.
XX Isolated protein binding domains from Plasmodium vivax and Plasmodium
XX falciparum erythrocyte binding proteins useful for vaccinating against
XX malaria.
XX Disclosure; Columns 79-92; 93pp; English.
XX
XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
XX (Duffy-binding like) gene family. The ebl-1 proteins are substantially
XX identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
XX Binding Protein (SABP), which are soluble proteins that appear in the
XX culture supernatant after erythrocytes infected with malaria release
XX merozoites. Immunochemical studies indicate that DABP and SABP are the
XX respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
XX and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
XX used to vaccinate against malaria, especially caused by P. falciparum.
XX Immunization with the polypeptide provides effective protection against
XX malaria. The present sequence represents the Proj3 binding domain
XX polypeptide.
XX
XX Sequence 2710 AA;
XX
XX Query Match 15.38; Score 2972.5; DB 21; Length 2710;
XX Best Local Similarity 26.56; Pred. No. 2.66-159;
XX Matches 865; Conservative 442; Mismatches 873; Indels 1083; Gaps 120;
XX
XX 9 IIKMNAASSLEGDAKSPITIKESHSAENVLEERYAKNIRHSK-YAKENHVSIGDLTKA 67
XX 6 LAKMGPKNAGGDIED-----ESAKHMFDRIGKDYDKYKBAKRGKGLGRSEA 58
XX
XX 68 EFRGSPSTPVNKHNYYPNCLDKERTNLRYDVNLRHPCHGREGONRPFDEDESEC-G 126
XX 59 KFEKNESEDQTPED-----PCDLLHKHTVY---TTVINPCADRSVRFDEYGGGCTH 110
XX
XX 127 NKIRYKKNKNAIACAPRRRHCKDKNEALNDINTONIHLLGNVLTATYBESLYNN 186
XX 111 NRIDSQGDGKNGACAPRRRLHVCDOLEQTEPIKTTHNMLVDVCAAFECOSITOD 170
XX
XX 187 HP-HKGT-----SDACTALASFPADIGDIVGIMKERNVND-----KVENGLREVFKI 235
XX 171 YPKYQATGDSPSQICHTLASFADIGDIVGRILYLGNPETIOROOLENNLTITIGKI 230
XX
XX 236 HD---GMEDEVKNDYNPDGSGNYKLRBAWNNVNRKNVKAITDASYSKGYEQSSESNT 292
XX 231 YEKINGAARFGND--PE-----FKLRMDWTANREYVWAKITGNA-WGNIFY-HATCNR 282
XX
XX 293 PLFSNPKCGHKGQVPTNLDYVPTLRFDEWGEFCRKNIKLKKYKDSR---NDKE 348
XX 283 GERTKGYCRCNDQVPTFYFDVPTLRFEEWADFCRKNKKIKDYKRNCRGKDEKID 342
XX
XX 349 RLYCSHNGHDTTITWKGILHLDNKCDCSTKCKVEVWLGNOEAFKKQKEVKEKE 408

```

RESULT 6
AA77904
ID AA77904 standard; Protein; 2710 AA.
XX AA77904;
XX
XX 13-JUN-2000 (first entry)
XX
XX P. falciparum Proj3 binding domain polypeptide.

Db 343 R-YCSNRGYDCEKTRAIKGRGKOCISCLYACNPVYDMINNOKEDFOKOKKYEIK 401
QY 409 SY-----LSNDKFEVNNINSE-YKQFYEKLEETOYATNDTELNLNECKYC----- 454
Db 402 KYENGASGSGSRQKRDAGGTTTNTYDGEKFEYDELNKESEYRVDKLEKLSNEICHTKV 461
QY 455 --KGLPGEEDITFTNSA-----DDKGFYSEYCOVCPDGGVCKDGIKTH-KSDND 504
Db 462 DEEGGTTDFKNVNSDSTSGASGTNSESQGTFFYSKCYQPCPYCGVK-----KVVNGSSNE 517
QY 505 REHANN-----EDYKPRMGVKTPTNITVYLSGNDGDIQOKLEFNC-----SS 547
Db 518 WEEKNNKCKSGKLYEPKPEKETTITILKSGHDDIEKLNKFKDEKNGDITNSGSG 577
QY 548 TNYKDKNN-----OKMECYRDENINRCKLEONTLEINNDNR-----IT----- 586
Db 578 TGSSGGSNSGRQELYEEMKC-YKGEDVYKVGHDEDEDEYENKNAGLICILKNOCKNKE 636
QY 587 -----SFHNFELMYTYLLDITIKNDKLTCT-NNTTTHC-IDECNRNC 629
Db 637 EGGNTSEKEDDEIOKTFNPFYVVAHMLDSTIHWKKILORLONGNRIRICGNKKNKND 696
QY 630 LCFDWWYKOKKEEMNSIKLFTKKNKNO-----OSYSNINNLFFGYF----- 672
Db 697 ECKRMTQKDKDMGKIYOHF-KTONIKGSGSDNTAELLIPDHDIYLOYNLOEELKGD 755
QY 673 -----FKYMDKLDEAKKELMENIKRKNESNLENNRD-----YLENAIELL 718
Db 756 SEDASEKSENSIDAEBE-----EELKHLREIIESEDNNQOASVGGVTEQKNIMDKL 809
QY 719 DILKERTATC-----KDNNTNEACETSHNATITNCPVPRGTO-PTKNIKELAQ 766
Db 810 NYKDEADLCLEIHEDEEKEKGDGNECTEESGNFYKNSGSGKRRPYLANKAYQ 869
QY 767 YFRSAVEEARNGLHLKKAHEGIYKRGRRKDFKDNLCRIMIKNSNLTGSPNGCP 826
Db 870 MHHKATQULASRAGSLARBDISLAQFKNNGNSTLKGQCKINENKSNDSRNGSGPCT 929
QY 827 GKGTGCIOTRFYVTEWEYDEPHMKRKHEDVMPRRRHICTSNLEHLOTDDHPLNGT 886
Db 930 GK-DGHHGVMRIGITWMSIEGKKQTSYKNVFLPRRHECTSNLEHNDLV-----GSV 982
QY 887 V-DOLVNSFLGVLNLSAKYEANKIITMYEKEKNLKGPRKVTDP---KHOTTICRATYS 942
Db 983 TKNDKASHSLGVDQLAAKDAEILIKRYKDNNT---OLTPLOOKDDEANCRAYRS 1038
QY 943 FADIGIIRGRDLMERNGDVAVKLOHLETFVGNIHKSLNG-KGNDKYND---APRYKL 998
Db 1039 FADIGIIRGRDLMERNGDVAVKLOHLETFVGNIHKSLNG-KGNDKYND---APRYKL 998
QY 999 RENWMEANRAKVAEMKCDIKYILKNSGHOSSTOSSYCGYSDHPLDDYIPOKLMMTEMA 1058
Db 1099 RADMWMEANRHOVRAMKC-----ATKGIIC---PCMHVDDYIPQRLMMTEMA 1143
QY 1059 EWTCKYOKKEEYDLKECKECKDKNDNGOGCTKSGCTCTACENKYNNDIIGLMEQWN 1118
Db 1144 EWTCKYOKKEEYDLKECKECKDKNDNGOGCTKSGCTCTACENKYNNDIIGLMEQWN 1201
QY 1119 IISDKYKLEHQAOMYSNSGIEASSTAKNHD---RNVLEFLSELQOONGK-----SNK 1171
Db 1202 KISDKYVLYLQAKTSTNG-----RTVLGDDDDPDVGOAMDPLPIHKAISAARVLYKRA 1257
QY 1172 SGTSDESASVIGTITTEVNCAYILHDGNPDQSONEFCOEK-----SPGKDNKYAFR 1225
Db 1258 AGSPTEIAAAPTTPYSTAGAYIHOEIGYGCQEOOTFCCKKHQAVSTSTYKKEKETYFK 1317
QY 1226 DKRODDHAGCGCKSGSPTAVOJITKKKAEKDEKTEVNDILKENDGKGOVEDCHPKN 1285
Db 1318 QPPEVATACDCTNRSOTE---EPKKKEENVESACIYKELIENKNGRTTVEGCNKES 1373
QY 1286 SNGPYMOC-GNINLVDEPRVCMPPRRQKLCVHFLANDNEIKKLOSOVNLEAFIKSAA 1344
Db 1374 ---YPMDCCKNNIDISHD-GACMPPRROKCLVYIAHESQENIKTODNLKDAVITKAAA 1429

QY 1345 ETEFSWYIYKSK-DGEGNELDKELKEGKIPPAFLRSNFTYFGDYRDLFGTDISKGGEG 1403
Db 1430 ETEFSWYIYKSKDNSEKAILDR-----GLIPSOFLRSNFTYFGDYRDLFGTDISKGGEG 1485
QY 1404 SKLEQDISLEFKNDKOSSPNGKTRQEWTEHSHETWAMCALVKAIAKDD---FTEHY 1460
Db 1486 AKAKDKIGKFFSKDGSFSGLSHQEWKKTNGPEIMGMCLALKRYTDTDNKRRIKNDY 1545
QY 1461 GYNNVKSDEK-SPTLEEFARPOPLRMWLEMYDGYCYTRQKYLADVOECKS-NDOLKCD 1518
Db 1546 SYDVNOSQNGNPSLEEFARPOPLRMWLEMYDGYCYTRQKYLADVOECKS-NDOLKCD 1605
QY 1519 T---EKNKCEDYVYKYN-KKEMIPQDKYXKODERKRRFDROHIGVATDYGTNATDY 1574
Db 1606 DAKHRCNOACRAVOEYENKKEKFSQGTNNFVLKANVQPODEYKGYEYKO----- 1656
QY 1575 LNKRTASCGDPRSASVQNRNIDLEKQAYIADAKRGCTKFTLENDKTNISKDKC 1654
Db 1657 -----GVQP-----IGNEYLQK---CDNNK-CSC----- 1678
QY 1635 GLVKEANTGAIKQONKGPNNVNNLKELEDEVLFPSSRLICFHALDGNITDPEYKDEGL 1694
Db 1679 -----MDGNLVSPEK----- 1691
QY 1695 RKRLMEVATEGYNLGOYKYEKKEKERTSDAHKYSYEVPPCSAMKYSFYDLRIDIGI 1754
Db 1692 -----PFGRY-----AHKYP----- 1701
QY 1755 DNLEDEKQTEENILKIFKNKNGTSVKGSDSTTGNPSTARKFFWNNKCVWNAIMCIGY 1814
Db 1702 ----- 1701
QY 1815 KRGHDSNGNSASRDEDLKKGCSVPDDDYPMKKNDEGTAQOFLRMFAWGEDPCKHK 1874
Db 1702 ----- 1701
QY 1875 EKLEKLVGACNDYTCGDNEDKRRKCTDCTOYKFTISEMPOYKOKIKYGENKXIYS 1934
Db 1702 -----EKC----- 1704
QY 1935 EHPAKDAEDAREYLDQKLIKCNKSGDCEYKCMKDYSTQRLDNGSQNMPSALDEPK 1994
Db 1705 -----DC----- 1706
QY 1995 EVEGCKCQVPRGRPRVRRPSPBVSLSKATASKKEAKTAPTKOPKVENLTTEMA 2054
Db 1707 -YQCK---HVPSP-----PPPPVQ----- 1725
QY 2055 QTRRAAQQTRKRTSTATTESDVGMVKAISLKNPDSRGIEGCKNPKTYGQY--PRWG 2112
Db 1726 -----PEAPVTVYDVCSIVTLFK---DTNNFSDACGLK-YGKTAPESSWK 1766
QY 2113 CI-----YKSKENNGICMPRRKRLCINNIOYNTET-----ENKRD 2151
Db 1767 CIPSDTSGAGATTGKSGSDGSCICPPRRRRRLVYGLQ--EWATYALPQEGAAPSRSRA 1824
QY 2152 NDKEAFIKCALETOFLMKYIIE-NPAEENE-----LNGT 2188
Db 1825 DDLRNATIOSALIEFTFLMBRYKREKRPQDGSQOALSQLSYSTSDEEDPDLDLNGK 1884
QY 2189 IPDEFKRIAMTYTYDYDMFFGDISNDKTIITYTNSVTTILNENKKNKODKKDEELRK 2248
Db 1885 IPPDFLRLMEYTYLDYDILVHGNGTSDSGNTGNSNNNIVLEASGNKEDMOKIOEKIEQ 1944
QY 2249 I-----FWEKNKFTMEGAYIGTY-----HLDENKKEIR 2280
Db 1945 ILPNNGSTPLVPKSSAQTPPDKMWEHAEISYKKGICALITTEKAPDTSARDEKKEIKED 2004
QY 2281 DNY-----QYN-DWTKL-----PSSLEEFVYKRP 2302
Db 2005 EYERKFFGSGTADKRGTASTPTGYTKQYDEKVLTEDTSGAKTPSASSDPILSDVFLRP 2064


```

QY 2303 QFLRMETMAEECKRRKREOLIKLEAGCKEYF-----CN---GSN 2339
Db 2065 PFRLEEMGQNFCKRRKRLKLAQIKHECKVEENGSGSRGRTIROYSGDGACNEMLPKN 2124
QY 2340 DKTQOE-----CAEACVITYONFIKKMTEYEROREKREKDD----- 2376
Db 2125 DQVVPDLKPPCAKPCSSYRKRIWIESKKEFEKEQAKAYQOKDKCVNSNKKHNGFCETLT 2184
QY 2377 -GKKKKY-----PSTERDIEKATCAHEYLMMKLKELGNGDCSCMOKPSSQLPRTQOS 2430
Db 2185 TSSKAKDLKTLGPCCKPNNVEGKTIFFDDKTFK-----HTMDCPCLKFSYNCKK----- 2234
QY 2431 QSSDAMPESLDYVPEEFNCECEPELSKSGSMHTKKTTPKIPMNCVEKAAVYLSKEA 2490
Db 2235 -----DECD-----NSKGTDCRMK-----NSID-----ATDI 2256
QY 2491 ENNDITLKEKFIPIESTKEKSNWTNNPCKPKPYAPDKYIGRRNCPEN----- 2533
Db 2257 ENGVDSTV-----LEMVVSADSKSGFNGD-----GLENNCRGAGIEGT 2295
QY 2544 REENRFYDYMCKYKSKFTYQEKKRVCPVPRRHMLRLNDELKIRLSDNYLTKMVR 2603
Db 2296 RKD-----EMKC-RVWCGTV-----VCKP----- 2313
QY 2604 RTARNEGIDITKNFNSENGAMNPICDTMKYSFADLDIVRGDMLRIGYLPPEVETKLY 2663
Db 2314 -----ENVNGE-----AKGHHIIOIRALVR-----RVW 2336
QY 2664 KVEFYIYKRRNKKNGKNNKYND-----VQTRSAMWMDNRKDWMKAMTCKAPEDAKLEFRG 2719
Db 2337 EYFEFEDNKIKRKHISHRIKNGEISPCIKNCVEKVMVQKRE--WKEITERKD----- 2387
QY 2720 RMDGFERITLIDKCGKHDPDYDITIPORFRMTEMSEYTCALMBELKEFEKSCDHCK 2779
Db 2388 -----QKKNDSDDNVRSFLETLIPQITDAN--ANKKIKLSKFGNSCG--CS 2432
QY 2780 TSDRCNKNDYDNCKECCQKTRCOEYKNFVLKWSLFIQSNKYELEYEQPIYKISYIVHY 2839
Db 2433 AS-----ANEONKN-----GEYKDAI-----DOMLKLKD-----KIGCECK 2465
QY 2840 QNFVQALKTFKSCVSESEYIYHETSCLNTKFNENDSSNTRYAFETPRSYKE--- 2896
Db 2466 HH-----QTSDECSPTPOPQULED---ETLDDIETEERAKNNMMPKICBNVLTQAQOE 2518
QY 2897 -ACSTLPSPKNPLDNCPTDOND 2918
Db 2519 GGC---VPAENSEEPALITDSKE 2538
UTL 7
MAR70236
ID AAR70236 standard; Protein; 2703 AA.
AC AAR70236;
XX
XX 22-SEP-1995 (first entry)
DE P. falciparum Proj3.
XX
XX Erythrocyte binding ligand; Proj3; binding domain; malaria; therapy;
XX
XX vaccine.
XX
XX Plasmodium falciparum.
XX
XX WO9507353-A.
XX
XX 16-MAR-1995.
XX
XX 07-SEP-1994; 94WO-US10230.
XX
XX 10-SEP-1993; 93US-0119677.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

```

```

XX Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI Wellens TE;
XX
XX WPI; 1995-123427/16.
DR N-PSDB; AA083529.
XX
XX New erythrocyte binding domain polypeptide(s) - isolated from
PT plasmodium binding proteins, used in diagnosis, treatment and
XX prevention of malaria
XX
XX Disclosure; Page 61-65; 81pp; English.
XX
XX Erythrocyte binding ligand (EBL) family genes were cloned from
CC P. falciparum chromosome 7 subsegment libraries constructed during
CC genetic studies of the chloroquine resistance locus. The 4 genes,
CC EBL-e1 (AA083526), EBL-e2 (AA083527), EBL-e3 (AA083528) and Proj3
CC (AA083529), encode the proteins given in AAR70233-36, respectively. The
CC binding domains of such proteins can be expressed e.g. in E. coli,
CC yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
CC cells, and provide protection against P. falciparum.
XX
XX Sequence 2703 AA;
SQ
Query Match 15.3%; Score 2972; DB 16; Length 2703;
Best Local Similarity 26.6%; Pred. No. 2.7e-169;
Matches 861; Conservative 438; Mismatches 864; Indels 1076; Gaps 119;
QY 33 KSRANVLERYANNIRHPSK-YAKEHVSLSKGLTFAEFGPSTPVNKHNYTPPCNLD 91
Db 16 ESKAHMFDRIGDYYDYKKEAKEKRGKGLQRLSEAKFEKNESDQPTED-----PCDLD 70
QY 92 HKEHNTLRDYNLHHPGHRQNRFDDESEEC-GKIRYKRRKNDALACAPRRRHMC 150
Db 71 HRYTHNV---TTVNLNCPADSDVRESDYEGGQCTHNRKISQOGDNKACAPYRLHVC 127
QY 151 DNLEALNDINTONIHDLGNVLYTAKYEGESIYNNH-RKGT-----SDACTLARSFA 204
Db 128 DONLEOIEPIKITNTHNLVDYCAAKREGOSITODIPKYATIGDPSQICTMLARSFA 187
QY 205 DIGDIVRDIIDFKPRVHD-----KVEGLREVEKKIHD--GMEDEYKNDYNPDGSGNY 256
Db 188 DIGDIVRDRDLYLGNPOIKORQOLENNLKTIPKIKYIEKLNGABARYGND--PE----F 241
QY 257 KLRAMNNVNNKRYWEALTEICDASYKSGYFQSESNITLFSMPKGGHOGAPYNTLDVPO 316
Db 242 KLRBDMWTANRETYWKALTCNA--WGNITYF--HATCNRBERIKGTCRCNDQDVPYTFDVPQ 299
QY 317 YLRWFDEWGEFCRRKRNILKLVKDSGR---NDKERLYCSHNGDCTTITMKGLILD 372
Db 300 YLRWFEEVAEDFCRRKNNKIKIDVARNCRGDKEDKDR--YCSBNGYDEKTKRAIGKLRYG 358
QY 373 NKCTDCSTKCKVFEFVWVWLGNOEAFKKOKEKEKEIOSY-----LSNDKFNANNIS 423
Db 359 KQICISCLYACNPYVDWLNNOEOPDKOKKRYDEIRKYEENGASGSGROKRDAGCTTTTNY 418
QY 424 E-YTKQFYERLKETQYATNDTFLNLNEGRYC-----KGLRGEKDITPTNSA----- 470
Db 419 DGYEKRRYDELNKSERYRTVDKFLKLSNEEICRYKRVYDEDEGTDIFRNVNSDSISGASGIN 478
QY 471 -DOKGIFYRSEYQVCPDGVKCDGIKTYH-KSDNDRERYNN-----EDYKPPMGVVKPT 522
Db 479 VESGTFYRSKYQCPCEYGVK---KVNNGSGSSNMEKNNKCKSGKLYERKPKDEGT 534
QY 523 NITVLYSGNEGDITOKLENFCA-----SSTNYKKDKNN-----QKMEYTYD 564
Db 535 TITILSKGKHDIIEKIKNFCEBERKNDIINSGGSGTGGSGGNSGRQELYEEMKC--YKG 593
QY 565 ENINRCKLEQNTETINDNPK-----IT-----SFHNFELM 595
Db 594 EDVYKVGHDEDEDEYENYKNAAGLCITLKNKKKKEEGNTSEKEPDEQIKTFNFFETIY 653
QY 596 VTYLNDITKIMNDLKTCTI--NNITTHC--IDECNRNCLCPDRVNVKQKEEEMNSIKKLITKK 653

```

Db 654 VAHMLKDSIMKKKLOKCLDNGNRKICGNKNCNDCECFKRWITQKDEKTYOHF-KT 712
 QY 654 KNIO-----QSYSNINNLFEJGF-----FKYMDLKDEAKKELM 690
 Db 713 QNKRGRGSDNTAELLIPFDHDIYLOVLOEELFKDSEDSAESEKSEMSLDAEFA----- 766
 QY 691 ENIKRKNEESNLENNND-----YLENAILLELDHLKETATIC-----KDN 731
 Db 767 EELKHLLEIESEDDNNOEASVGGVGTQKIMMDKLINYEKDEADLCEIHEDEEEKK 826
 QY 732 NTEACETSHNATNPNVCRGTQ-PTKNIKELAOYFKRSAYEARNRGIHKLKGKANE 790
 Db 827 DGNCEIEGENEFKYNPGSGSGKNRYVLANKAYQHNNHAKTQULASRAGSALRGDLSL 886
 QY 791 GIYKRGRRRDFDNLCRIMIKHSNRMJGFSNCPDGTGDIOTFFVVTGWEYVPEH 850
 Db 887 AOFKNGNGSLTQOIKIKIMENYSNDRSGSGGCTGK-DGDHGVGMRRIGTSEMSNEGK 945
 QY 851 MKKDHEVIMPERRRHICTSNLEHLQTDHPLNGNIV-DDLVNNSLGADVLLSAKYBANK 909
 Db 946 KQSYKKNVFLPBRREHCTSNLENDY-----GSVTKKNDKASHSLGADVOLAKTDAE 999
 QY 910 IIRMYKKNLKGKKEVTD----KHOTICRAIFRSADIGDIRGDLMEERNNDMYKLO 966
 Db 1000 IIRYKQONNI-----QLTDPLOQDOEAMCRAVYSFADIGDIRGDMDEKSTJME 1055
 QY 967 GHLETVGNHHSKLG-KGDKNDYND---APKYLKRENNWEANRAVWEAMKCDIYK 1022
 Db 1056 TRITTVKKNKEKHNDGKIDKPNKTYGDSKRPAYKKLADWEANRHOVYRAMKC----- 1109
 QY 1023 DKSHOSTOSYSGYSHTPLDIYIPOKLMMTEMAEMYCCKVQKKEYDKLKECKECKD 1082
 Db 1110 -----ATKILIC---GMPADYDIIPQRLMTEMAEMYCQAOQBEPDKLKCICADCMK 1160
 QY 1083 DNGOGCTKESGTGCTEACNEYNNDIIGLMEKDOMNIIISOKYELHROAOQMSVNSGIEA 1142
 Db 1161 GOGK-CT-QGDVDCGKKAACDKKELEKNEWMKRIISDKTINILYLOAKTISNPG--- 1215
 QY 1143 SPTAKNHID---RNVIEFLSELYOONGK---SNKSGTDESNAVIGTNTTYENVGAYLH 1195
 Db 1216 -RTVLGDDDDPYQOMWDFLPIHKSIAARVLYKRAAGSPLEIAAARPIPIYSYACIYH 1274
 QY 1196 DTGDFDCQSONERCDK-----SDCKNEKYAFROKPODHGACGCKSGSKRTYQIK 1249
 Db 1275 QELGYGCOBOTOPECKKHATSTYTKENKEYFFKQPPPEYARACDCINSQTE---E 1330
 QY 1250 TKKAEEKDECKTVNDILKENDGKQOYEOCHPRKNSNGYPDQOC-GNINLYEPRYCM 1308
 Db 1331 PKKKEEVESACKIYERILBKKNKRTVGECPKES---IPDMCKKNNDIDISHD-GACMP 1386
 QY 1309 PRROKLCVHFLANDNEIKLOSOVNLKAEATKSAAEFFSWYYSKSK-DEGENELDEL 1367
 Db 1387 PRROKLCIYIAHESQENIKTDNDLADAFIKTAAEFLSMOYYSKSKNSEAKITLR- 1444
 QY 1368 KEGITPAFLRSMYTTGDRDPLFGDISKGHSGSKLEQIDSLKRNQDQSPNGKTR 1427
 Db 1445 --GLIPQFLRSMYTTGDRDPLCLNDLSDIKSKONDVAKADIKGFKPSKDSKPSGISR 1502
 QY 1428 QEWMTESHEIWEAMLCALYVIGAKKDD---FTENYGINNVKFSDK-STILEEAKRPOF 1483
 Db 1503 QEMMKTKGPEIKMCLCALIKRYTDTDNKKRIKNDYIDYKVNQSONGNSPISLEEPAKPOF 1562
 QY 1484 LRMWLEMYDYCYTRORYLADVDECKS-NDOLKCDT---ECNKRCEBYVYKAM-KRKEM 1538
 Db 1563 LRMWIEGEEFCARCKENIKKDCAGNEINSTOQCNDAKHRCNQCACAYOYEVNKKKEF 1622
 QY 1539 IPQDKYKDERDKRPRROHIGVAVTYDTGNAIDYLNKRTYASCGKPSGASVYQRIQ 1588
 Db 1623 SGQTNNVFLANVOPOPEYKGYEKD-----GVOP-----IQENEX 1659
 QY 1599 LLEKQAYYDADKHGCTKFTENDOKTYNISKDKGLVKEANTGAIKMONKGPNNNTNL 1658

Db 1660 LLQK---CDNKK-CSC----- 1671
 QY 1659 KELTEDVLFPSRRRLICFHALDGNYPDEYVDENGELKRLMEVAATEGYNLQYKKEKE 1718
 Db 1672 -----MDGNVLSVPREK-----PRGR----- 1689
 QY 1719 KETIKTDARKYSTEYVPCSAKYSFYDLARDIILGIDNLEDEKOTEBNLKIFNNKSTS 1778
 Db 1690 -----AHKYP----- 1694
 QY 1779 VKGSDSTGQPGSTARKFFENNEKBEVWNAMICGYKRGDRDGSNSARSDELLKCGS 1838
 Db 1695 ----- 1694
 QY 1839 VPSDDDYPMGNRDEGTAYOLFMRFAEMGEDFCHEKELEKLYGACNDYTCGDNEDRK 1898
 Db 1695 -----E.1695
 QY 1899 KCTDAGTYKKFISEMKPYQEKQIKYGENKQKITYSEHPVAKDAEDAREYLOLKICE 1958
 Db 1696 KC----- 1697
 QY 1959 NKSQDCEYKCMKQVSTQRLTDGNSQNMPSASLDEPEKEGCONCQVGRPRVARETSP 2018
 Db 1698 ---DC-----YQK---HVSIP- 1710
 QY 2019 RVSLISKATASKKEAKTAPPTQOPKRYENLTETEMRAOTRTTAAQOTRKRTSTATTESD 2078
 Db 1711 -----PPPVYQPO-----PEAPTYVD 1727
 QY 2079 VGINVKAILSKPSPRSRGIEGECNPKTYQY-PRWGT-----VKSKEENGT 2125
 Db 1728 VCSIVTKTLF---DTNNSDAGLK-YGTAPSSMKCIPSDTKSAGATYKSSSDSGSI 1783
 QY 2126 CMPRRKRLCINNOYANET-----ENKRDNDIKAFIKCAIEFOFLMKYII 2175
 Db 1784 CIPRRRLRYGKQ--EMATLALQGBGSAAPSHSRADDLNRAFIQSAIETFFILMDRYKE 1841
 QY 2176 E-NPAENE-----LONGTIDPEKRIYVYVYGDYKDMFGTD 2212
 Db 1842 EKKPOGDSOQALSQLNSTYSDEDEPPDKLQNGKIPPELRLMFTYLDGYRDLIVHGG 1901
 QY 2213 ISNDKIIITYNSTTILNENKKKODKKDEELKI-----FWE 2232
 Db 1902 NTSQSGNTGNSNNNVIYLEASGNKEDQOKIOEKIEOILPKNGTPIVPKSSAOTPDKWN 1961
 QY 2253 KKKFIEGMYGLTY-----HLDENEKEXIRDNV----- 2283
 Db 1962 EHAESIMKICALITYTEKNPDTARGBDENKIEKDEYTERFFSGSTADKGTASTPTGTY 2021
 QY 2284 --QYN-DMTKL-----TPSLEEFVVRPOFLRMFTEMAERFCNKRREOLIKL 2326
 Db 2022 KTYDYVEKVLIEDTSGAKTPSASDTPLLSDFVLRPPEFRVLEMGOFCKRRHKLQI 2081
 QY 2327 EAGCKEYB-----CN-----GSNDKTOE---CAECVITYONFIK 2358
 Db 2082 KHECKVEENGSGSRGGITROYSGDGEACNEMLRKNDGTVPDLKPCSAKCSYSRWIE 2141
 QY 2359 KWKTEYEROREKFKKDD-----GKKYKDY-----PETERIDEAT 2394
 Db 2142 SKGKEFEKOEAYOQADKCYNGSNKNDNGRCETLITSSRAKDLTKLGPCKPNNVEGKT 2201
 QY 2395 CAHEYLNKKIKELCGNDCSCMQRPSSOLPRTTQOOSOSDANDMPESLDVYVEEFNCEC 2454
 Db 2202 IFDDOKTFK-----HFKDCDCLKFSVCK-----DECD- 2231
 QY 2455 PELSCKGSMITHKITYEPKIPMNCVEKRAAYLSKEAENNDITLKEKFIPIESTKEKESK 2514
 Db 2232 ---NSKQTPDCRNK-----NSID-----AUDIENGVDSTV-----LEMVSADSK 2267
 QY 2515 NSMTNNNPNCDPKRPYADKYIGRRNPECN-----REENFQVDYEMKCYKNSKPYQEK 2567
 Db 2268 SGFNGD-----GLENNACRGAGIFEGIKRD-----EMKC-RNVGTY-- 2302

RESULT 8	
AA18352	AA18352 standard; Protein; 2197 AA.
XX	
XX	AA18352;
AC	
XX	
DT	07-NOV-2000 (first entry)
XX	
DE	Plasmodium falciparum chromosome 2 related protein SEQ ID NO:210.
XX	
KW	Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW	antimalarial; malaria; protozoacide; infection; insecticide.
OS	
XX	Plasmodium falciparum.
PN	MO200025728-A2.
PD	
XX	11-MAY-2000.
XX	
XX	05-NOV-1999; 99WO-0526796.
PR	
XX	05-NOV-1998; 98US-0107131.
XX	
PA	(HOEF/) HOFFMAN S.
PA	(CARU/) CARUCCI D.
PA	(GARD/) GARDNER M.
XX	(VENT/) VENTER J C.
PI	
PI	Hoffman S, Carucci D, Gardner M, Venter JC;
DR	WPI; 2000-365347/31.
XX	
XX	
PT	Proteins encoded by chromosome 2 of the human malarial parasite,
PT	Plasmodium falciparum, useful as antimalarial vaccines and in the
PF	diagnosis of P.falciparum infection -
XX	
PS	Disclosure; Page 441-447; 577pp; English.
XX	
XX	
CC	The present invention describes proteins and their fragments (I) encoded
CC	by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC	Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC	vaccines against P. falciparum infection comprising (I) or (II).
CC	(I) and (II) are useful for the development of vaccines against
CC	P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC	antibody raised to immunogens comprising the sequences of (I), are
CC	useful in the detection of infection with P. falciparum. Furthermore,
CC	(I) (especially when they are filins or secreted or membrane proteins)

[illegible]

Db 743 SADVLELAL-----ELEOLF 758
QY 769 KRSAYEARNRGLHLKGAHEGIYKRGGRKDEKDLICRIMIKHSNRNLGFSNGPCDG 828
Db 759 -----DIIKO----- 763
QY 829 GTGDGIGTFRVVGTEWEVDEEHMRKDEHYIMPRRRRHICTSLNLEHLOTDDHPLNGNIYD 888
Db 764 GYGD-----VKELKGIKELDEBKKEKQAEAVV-----VV- 794
QY 889 DLVNSFLGDVLLSAKYEANKIIRMYKEKNLNGPKVEYDHPKHOTTCIRAIRYSEADIGD 948
Db 795 -----VADNOKKTYTIDKLQHE----- 811
QY 949 IIRGRDLMERNGDMVXLOGLHLETFVGNHSLKGNKNDKYNDAPKYLKIRENWEANRA 1008
Db 812 -----GDDANNCLKTH----- 822
QY 1009 KWEAMAKCDIKYLDKDSGHSOSTOSYCGYSDHTPLDDYIPQKLRMTWMAEWYCKVQOKE 1068
Db 823 ----- 822
QY 1069 YDLKREKCEKCDKD--NGOGCTKESGCTGCTACNEYNIDILGLMKEQWNIISDKY 1125
Db 823 -----KEKCEGTQPRPGAGPGAPSETGETT-----LEDEEE 856
QY 1126 ELHEQOMSVNSGIEASSTAKNHIDRNVIEFLSELYOQNGKSNKSGTDESAVIGTNT 1185
Db 857 EEDEE-----EDAGDEV-----EEGETVDT- 876
QY 1186 TYENVCAYLHDGTFDCCOSONEPCDEKSPDKONEKAYAPDKPODHDGACGCKSGSKPTR 1245
Db 877 -----TEGDETEVEQPVKDTDREG----- 896
QY 1246 VOJTKTKKAEKDE--CKTYNDILKENDGKKQVEDCHPKKNSNG--YPMQC--GN 1296
Db 897 -EEBEAKKATDTTSLDYCTVKNALTNND--NLTDACKLKYGBGGERPPMKCVSSGE 953
QY 1297 INLV-----EDPRVCMPPRRROKLVHFLANDNEIKKLOSOVNLKEAFIKSAA 1344
Db 954 KSVATGSSGATGSGDKGALCVPRRRRLVYGLTK-----LTSAGT 996
QY 1345 ETFFSMYYKSKDGEKNELDEKELKEGKIPPAFLRSMFYTG DYRDLFGDISKHGEGS 1404
Db 997 SS-----ESPQG-GSESSR-----ASDVSOQNG-- 1018
QY 1405 KLEKQJDSLEKNGDQKSPNKTRQEWMTESHSHEIWEAMLCAVLKIGAKKDDFENYGYNN 1464
Db 1019 ----- 1018
QY 1465 VKESDSTLEBEAKRPOFLRWLTWEYDYCYTROKYLKDOECKSNDOLKCDTECNK 1524
Db 1019 ---GDITTTESLRK-----WFE-----TAA 1097
QY 1525 CEYVYKMKKKKEWIPQDKYKDEBRKKRPDRQIHGMVMTDYGTNATDYLNKRTASCG 1584
Db 1038 IETFFLMHRYKKWEAKK----- 1056
QY 1585 DKFGSAYVORNIQLEKQAYYADADKHGCTKFIENDDKYTINISSKDKGVLKEANTGA 1644
Db 1057 -----AELORNGLL-----GTGA 1070
QY 1645 IKWONKGPNNYNNLKELTEDEVLEPSSRLRICFHALDGNYPDPEVK--DENGIRKRLME 1700
Db 1071 S--LNLGGDOSNPOTLOKSGTIPDLRLMFEYTL--GDYDILVRGBVADDKNG----- 1120
QY 1701 VAATEGNLGOYYKEKKEKIKTSDAHKYSYEVPPCSAMKYSYDLDRIILIGDNLDE 1760
Db 1121 ---GNN-----TILNAGNKDE 1134
QY 1761 KORTE---ENLKIIFNKGTSVKGSDDSTGNGSTARKFEMNENKCVNNAMIC--YK 1815
Db 1135 KOMEKIQEKIEQILPTISGKKEKTRGPONSVD-----RQSLMDRIAHVHWGHVCAALTYK 1189

QY 1816 RGRDNGSGNSANSDEDLKCGSVSPDDDYPMGNRDEGTAYOFLRMFAEMGEDFCKHKE 1875
Db 1190 ---DDONG----- 1194
QY 1876 KELEKLVGACNDYTCGDNEDKRRKCTDCTQYKKFISEMFPYEOIKKYGENKDKIYSE 1935
Db 1195 ----- 1194
QY 1936 HPAKQADAREYLDQOLKIKICEKSGDCEYKCMKQVYSTQRLDNGSONNPASLDDEPKE 1995
Db 1195 -----LKGV----- 1198
QY 1996 VEGKCNQVGRPPVRVRETPSPRVSLSKATASKKEAKTAPPTQPKVYENLTENRAQ 2055
Db 1199 -----YKPKQXEN----- 1207
QY 2056 TTRRRAAQTRKRRTATTTESDVGTVKAILSKPDSRGIEGNCNPTYGQYPRKWCIV 2115
Db 1208 ----- 1207
QY 2116 GKSEKENENGICMPPRRKKLCINNIOYLVNTEENKRONDIKAFIKCAIEFQFLMKYII 2175
Db 1208 -----PEK-----LM----- 1212
QY 2176 ENPAENELONGTIPDEKFRIMYYTG DYKDMFETDISDKKITVNSVTLINENK 2235
Db 1213 -----NETTK 1217
QY 2236 KQDKKADDEELRKIFWEKKNKFIVEGMATYGLYHLDENEKEKIRONYQYNDMTKLTPLSL 2295
Db 1218 KPRDEKYOYQAKL-----EDESERGEKRPDSASAG--TKLT-- 1250
QY 2296 EEFYKRPQFLRMFEWMAEFCENKREKQDLKLEAGCKRYEONGSNDGTQCAE----- 2348
Db 1251 -DPIKRPYRYLEWENFCCKRTMIGKIKEDC--YKNGRCSGGLKCNELVIDDEK 1307
QY 2349 -----ACVYONFIRKKKRTEXEROREKFKDKDKKKYKDYSTERDIKATCAHEYL 2400
Db 1308 IFGDLPCPCARICRFYKKW-----INTKRD----- 1333
QY 2401 NMLKELCGKDKSCMCKPSSOLPKTTQOSSDANDMPSLDYVPEFPNKCCEPELSK 2460
Db 1334 -----EFNK----- 1337
QY 2461 GSMHTKITEPKIPNMCVEKAAYLSKEAENNDITLKEKFIPIESTKESKNSFTNN 2520
Db 1338 -----QSNAYSQKKKY-----BEENDSAOKNN 1360
QY 2521 NPCDPKRPYAPDKYGR--RNPCENRENEKDYEMKCYKNSKFYOEKKRVCPVPRRH 2578
Db 1361 GVGCTLKDDAAE--FLNLLKNGPCNKESENNKKADELDFKPPDPTFFDADN--CKP----- 1413
QY 2579 MCLRLNDEIKIERKDSNYLLKMYRTARNEGIDIIENFSENCCAMPICD--TMKYSFA 2637
Db 1414 -----CSEFKIK-----CENHSCSSGNTQGRCDGTTIAAT 1445
QY 2638 DLGDIYRGT---DMLRIGYLPVEIKLYKVEYIYKWRKKNKGRKRYNDYQTFRSAMW 2694
Db 1446 EIEIKKTNTKEVYMLVSDSKSATEFK-----DGLSECKDGIFKGI-- 1487
QY 2695 DANRDIWKAMTCAPADAKLFRKGRMDGERITL--IQDKCGHKDDPVPVDYIPQPRRM 2753
Db 1488 ---RKDEMEC-----GKYCGYDICMLKKKNDIGRESD--KKYIIMK--ELL 1526
QY 2754 TEMSEYCKALMELEKFEKKSCHCKTSDCKANDYDNKBOCKTRQOEKKNFLYKXSL 2813
Db 1527 KRWLEYF---LEDYNNIKIKRISHCTRNK--GSKC-----IKGCYDKM-- 1564
QY 2814 FDIOSNKKELYEQPIYTKISTYDVONFQKLTFRKSECSVESSEYLAHETSKCLMYKF 2873
Db 1565 -----YQKKKEMKQIK--EFNNE-----QYKS 1585

QY 442 DFLNLLNECKYCKG---GLPBEKIDITFTNSADDKGIFRSEYCOVCPCCYKCDGICY 497
 Db 461 DKFLKILNKSGIOQOKPOVNEKADNVDFNEKYVK--TFESRTEICEPCPCGKEGPPW 519
 QY 498 THESDNDREVNNEDEYRPPGAVKPTNTIVLVSNEGOODIKOLENPNCSNSTNFKDKNNK 557
 Db 520 KYVGDKTCGSAKRTYDPK---NITDIPVLYPKDSOONILAKYKNCCEKAP--GGGQIKR 575
 QY 558 WECYKAD---ENINCKLEONTEINNDPKIISFHEFELWVYLLADTIRKMDKLKT 612
 Db 576 WOCYDEHRRSSKNNNKCVGTWDKFQOKQVKSYPVFEFMDVHDLHDSVEKTELSK 635
 QY 613 CINTTT---HCIDECNRNCLCPDRKVKKEEEMNSIKLFTKKNKIOOSYISNINLP 668
 Db 636 CINNNTNGMTCRRNNKCKTCCGCFORKEVKEQOEEMAIKJHFKQDIDYQ---KGLI 690
 QY 669 EGYEFKMDKLDKDEAKMKELMENIKRKKNEFSNLENNRYLE-----NAI 714
 Db 691 VESPYGVLDV---LKGNLONIKDVGDTDDIKIKILDEBDAVAVALGKDNNTT 746
 QY 715 ELLDLKETATICKDNNTNEACETSHNATNFCVPRGTOPTKNIKEIAQYKRSAYE 774
 Db 747 DKLOHKEBAEQCKK--QEBC-----KKAQ 773
 QY 775 EARNRGLHLKGAHAGCYKRGRRKDFKDNLCRIMKHSNRLGFSNGPCDKGTGDI 834
 Db 774 ESRGRSAE---TREDER-----TQPADSAGE--- 797
 QY 835 QTFVVTGTEWEDPEHMRKHEDVIMPRRRHICTSNLEHLOTDDHNLNINYYDLVYNS 894
 Db 798 ---VEEEDDDYDEDED---DVVQEE 820
 QY 895 FLGDVLLSAKYEANKIIRMYKEKNLKGPEVTPDKHOTTICRAIRYSFADIGDIIRGD 954
 Db 821 EEG---KEBGTVTEVEVEVEET----- 842
 QY 955 LMERNGDMYKLOHLETVFSGNIHKSILKKGKNDXYNDAPRYLLKIREMWEANAKYWEAM 1014
 Db 843 --- 842
 QY 1015 KCDIKYKDKSGHSTQSSYCGYSDHTPLDDYIPQKLRMTWMAEWYCKVOKKEYDLKE 1074
 Db 843 ---VTE-----OE 847
 QY 1075 KCCECKDRNGOGCTRESGTCTCTEACNEVNDIGLMEQWNIISDKYKELHEQAOMS 1134
 Db 848 GVRPC---DIYKGL----- 858
 QY 1135 VNSGIEASTAKNHNIDRNVIEFLSELVQONGKSKNSGTSDBSAVIGTNTYENNGAYL 1194
 Db 859 --- 858
 QY 1195 HOTGNFDDOSQONFPCDEKSDGKDNKEYAFRDKPRQDHGACGKSGSKPTRVOIKTKKKA 1254
 Db 859 ---FEDKSLKE---ACGLKYG- 875
 QY 1255 EERKOTBEKTYNDILKENDGKROYEDCHPKRNSNGYRPMOCGINLV-----EDPRVCM 1308
 Db 876 ---GGEK-----FRMKCVTPSGVSTANSBGDAICVP 906
 QY 1309 PRROKLCVHFLA---NDNEIKKLOSQV-----NLKEAFIKSAAEFFPSWY 1353
 Db 907 PRRRRLYVGLSOWASRGSEJTEVSSEASAPQSESEKLTAFISAAIEFEFLHKKY 966
 QY 1354 K---SKDG---EGNELDKELKE--GKIPAPLRSFYEFYGOVBDLFG 1393
 Db 967 KEKKRPATPDGAGLVSLEPSPRGEDPOTLOQOTGVIPFDLROMFTLADIKDL- 1024
 QY 1394 TDISKHGEESKLEQIDSLFKNGDKSPNGKTRQEWMTESHEIWEAMCALVYIGAKR 1453
 Db 1025 --- 1024

QY 1454 DDFTEYNGYNNVKSFSKSTTLEEFKAPQFLRMLEWYDDYCTTRQKYLKADVOECKSND 1513
 Db 1025 --- 1024
 QY 1514 QLKODTECNKCEDYVYKMKKKKEMIPQDYKYDERDKRRFRQHGIGVMTDTGNATD 1573
 Db 1025 ---YSGNDTS 1032
 QY 1574 YLNRKFTASCGDRPGSASVVOIRNIOLEKQAYYDADKHCCTREIENDRYTNISKDK 1633
 Db 1033 ---DTGQOTPESSNDL 1047
 QY 1634 KGLYKEANTGAIRKQNGPNYNNKLELTDVLEFPSRLAICFHALDANTDPEVKGNG 1693
 Db 1048 KNYLEAS--- 1055
 QY 1694 LRRLMEVATEGYNLQYKKEKEREKITSDAHKSYEVPSPCASAKSYFYDLRDIILG 1753
 Db 1056 ---GSTE---OEKKK- 1066
 QY 1754 IDNLEDEKOKTEENLKRIFNKNGTSVGKSDSTGNPSTARKFFMNEKCYVAMNIGC 1813
 Db 1067 ---QIOAKIKKILN-GATS---GVPPYTKNSVKTPOQTWMENIAKDITMANAVCA 1113
 QY 1814 --YKRGHDGNSGNSARSDEDLKKCGSVPSDDDYPMGKNRDEGTAYOFLRMPAEMGEDFC 1871
 Db 1114 LTYKE---NDAR-----GTS 1125
 QY 1872 KHKEKELKLVGACNDVTCGDNEDKRRKCTDAGCYKKTISEMKPOYKOKIKYGENKDK 1931
 Db 1126 ---AKIOKND- 1133
 QY 1932 IYSEHPYAKDAEDAREYLDOLKRICENKSGDEYKCMKDVSTOGLTNGSONMPASILD 1991
 Db 1134 ---LKK-----ALWDEANKNTP 1147
 QY 1992 EPKEVEBCKCQVPRGPRVRRRETSPRVLISKATSKBEAKTAPPTOKPKVENLTTE 2051
 Db 1148 --- 1147
 QY 2052 MRAQTRTARAQOTRRRTSTATTESDVGTMVKAILSNKPDNRGIEGNCNPTYGYPKW 2111
 Db 1148 --- 1147
 QY 2112 GCIVGSKENENGICMPRRRKLICINNIOYLANETENKROUNDIKAEAFKCAIETOPLML 2171
 Db 1148 ---IEKYQTYNVALE----- 1159
 QY 2172 KYIIEPAENELONGTIPDEFKRIMTYTGDIKDMFEFGDISNDKIIITVNSVTIILN 2231
 Db 1160 ---D 1160
 QY 2232 ENNKKQDKKDEELRIIFMEKNKKFIWEGMIYGLVHLNDENKEKIRDNQYNDMTKL 2291
 Db 1161 ESGAKSND---TIO 1171
 QY 2292 TPSLEEFKPPOLFRMTWMAEFCNKRKELKLEAGC-----KEYECNGSNGDK- 2342
 Db 1172 PPTLKNEVEIPTFRMLHENGNSCFERARAKIQAIRKECHDEDOEKQY---SGDGYCE 1227
 QY 2343 ---TOEBAECVYQONFIKKKWTIEBQREKFKKDKQKQKYYDYPSTERD 2389
 Db 1228 EIFSROYNVLODLSSCACPKCRLYKTIERKKEYEKQOAKAYEQK---SNYENOKD 1282
 QY 2390 IERATCAHEYLNMKLKELCGNKDCSQKPSQSLPKTQOSQSDADMPESLDYVP--E 2447
 Db 1283 ---KC-----OTOSNNANDEFKRTIGASPTAA 1306
 QY 2448 EFNKCEPELSKSGSMITHKRTIEPKIPANCVERAAVYLSKEAMNNNDITLKEKFTIPES 2507
 Db 1307 EF---LQKLS---CKNDNGY-----ENGEDNKIDFK-NPDKT 1337
 QY 2508 TKEKESKNSWNTNNNPPCDPKPYADKTYIGRRNCPENNEENRKFVYDWMKYCKNSKPYQEK 2567


```

Db 1338 FKEAHS-----CDP-CPITGVK-----CQNGH-----1358
QY 2568 KRVCPVRRBHEMLRNDEIKIERLSDSNYLKRVRTARNEGIDIKNFNSGCMANP 2627
Db 1359 ---CVGSANGKECKNN---KITAEDIKN-----KTDNGNIENAVSDSTN-----1398
QY 2628 ICDTKAKSPADLDIYVGTMLRIGLYLPEVEIKLYKVFYIGKWRNKNKGRNRYDVQ 2687
Db 1399 -----TFEHLGD-----CKSSGIFKGI-----1415
QY 2688 TFSAMWADRKRDIWK-AMTCAKPEDAKLFRKRMGFEITLID- KCHKDDPPVDY 2745
Db 1416 -----KDEMKCANVC-----GVDICTLEKIKIKNGQED---KKY 1447
QY 2746 IPRFRMTWSESYCCALMEDEKFKKSCDH- KTSDRCKNDYENKCBQCKRCQERY 2804
Db 1448 ITWK-ELLKRWLEVF-----LEDYNRIKKIKLCTKKEDGK-----CIKGLIE- 1490
QY 2805 NEFLKMSLEFDIOSNNKYKEIYEQPIYTKISTYHVQNFVOKLTKERSECSVESFSEYLYH 2864
Db 1491 ---KM-----VO-----EKTQWO-----1501
QY 2865 TSCKLYKFNENDSSNIRTYAFETPKSYKACSCTLPSKPNLDNCPTDQNDGCKEIQ 2924
Db 1502 -----KIND-----TY-----LE 1509
QY 2925 TTFPCSKNDYNDLDMWNAVLVNSSDNKGVLIPRRRRHLCTPITAYNKRKDEILK 2984
Db 1510 QY---KNDGNTLTF-----LEQOYR-----1529
QY 2985 KLLTSAFSOGQLGOKYKSEELCFEAMKSYADYSDIIGKDDMDTSLSEKIKIET 3044
Db 1530 -----TEFKNAIKQDGLD-----1543
QY 3045 SNEATENKRTWMENNRROIWHAMLGKYIATSKVTLDEGWCQLPKDETNOFLRWLIEMA 3104
Db 1544 -----QE-----1545
QY 3105 KOACKERKHVSDSLTKCPRSNEDNFEASELLRQPCCONDIRKYSILNLIKNTBENLNI 3164
Db 1546 -----KTSCLNSTDN-----SQNGNND-----LVLCILNKLQK 1575
QY 3165 KTKQKLDSSGINDKPSSEENVQYIKSDSOCALELNDINEIVGTRKNENN-----3217
Db 1576 KISECKEQHSQGTQF-PCD---NSSLSGKESFLVEDVDYEE---QNEPKVQQPKFC 1626
QY 3218 -FEKEVAKLYLGLYFVEDETHKNVLDGNT--KEEBQTVRPKALYFTTPHYDSFYQAPL 3274
Db 1627 PDKMEPKKENDSEVGTGCGDEEKKEVDSVLEQKEEEMASAPESPLTP-----EAPK 1680
QY 3275 FSTHVAQYDP-----KN-----DILKGSISVIVSALGL-----IALHFMKKRKS 3316
Db 1681 KEENVVPPRPPPKRRRIKTRNVLDHPAVITPALMSSTIMWSIGIFAAFTFYFLKKRKS 1740
QY 3317 SV-DLLRLINLPQGEYGMPTLESKNRYIPYRSGPYKGYIYMEGDSIGDEKYYMWDSS 3375
Db 1741 SVGNLFOLIOIPKSDYDIPITLKSSNRKYPYASDRHKGYIYMEGDSGDE- KAFKMDT 1799
QY 3376 SDTSESEYEELNDIYVGSPPYKTLIENVL-EPKRDIPSD-----DTSNDTPR 3428
Db 1800 TDTSESEYEELNDIYVGSPPYKTLIENVL-EPKRDIPSD-----DTSNDTPR 1859
QY 3429 TNRFIDEMNELKHFVSQYL---PTEPPNNNYKADIPMNTPEPTLYSDNPEEPFIS 3485
Db 1860 P--ITJDEMNQLKDEFTSNMLQNTQNEPNTLHDNV- ANTHP-TMSRHMDOKPFIMS 1914
QY 3486 IHDNDLYIGKELSY-----NINNSTN-----TNNDIPMANARSDRYGIDLINDS 3529
Db 1915 IHDRLKESGEYANVDMFNSGNPNINISDSTNSMDSITSNNHSPYNDKNDLVSIGIDLINDA 1974
QY 3530 LVVNLNLIY--MMK 3541

```

```

Db 1975 LSGNHIDIYDEMUK 1988
RESULT 10
AA77906
ID AA77906 standard; Protein; 2182 AA.
XX
AC AA77906;
XX
DT 13-JUN-2000 (first entry)
XX
DE Plasmodium var-1 polypeptide.
XX
KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KW DABP; Stialic Acid Binding Protein; SABB; malaria; vaccine; immunisation;
KW protozoacide; var-1.
XX
OS Plasmodium sp.
XX
PN US593827-A.
XX
PD 30-NOV-1999.
XX
PF 07-JUN-1995; 95US-0487826.
XX
PR 10-SEP-1993; 93US-0119677.
XX
PI Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;
XX WPI: 2000-194198/17.
XX DR N-PSDB; AAZ98288.
XX
PT Isolated protein binding domains from Plasmodium vivax and Plasmodium
PT falciparum erythrocyte binding proteins useful for vaccinating against
PT malaria.
XX
PS Disclosure; Columns 129-140; 93pp; English.
XX
CC The invention relates to ebl-1 polypeptides that are encoded by the DBL
CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
CC identical to the Duffy Antigen Binding Protein (DABP) and Stialic Acid
CC Binding Protein (SABB), which are soluble proteins that appear in the
CC culture supernatant after erythrocytes infected with malaria release
CC merozoites. Immunochemical studies indicate that DABP and SABB are the
CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
CC and stialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
CC used to vaccinate against malaria, especially caused by P. falciparum.
CC Immunization with the polypeptide provides effective protection against
CC malaria. The present sequence represents the var-1 polypeptide.
XX
SQ Sequence 2182 AA:
XX
Query Match 9.0%; Score 1743; DB 21; Length 2182;
Best Local Similarity 20.3%; Pired. No. 1.9e-95;
Matches 757; Conservative 339; Mismatches 678; Indels 1960; Gaps 126;
QY 13 GNAASLEGDAKSPILNESHKS-ARNVLER-----YAKNIRHPSYAKEHVDLSKGDLYK 66
Db 10 GSGGSS-SGKGGKDTSEYIYVSDAKDLDRVGEKYVEEKYKND--AKKTIEMKGNLNT 66
QY 67 AEPFGSGSTYVKNHNYPPPCNDLHKEHTNLRYDVN---LHPCH---GREQNRPE 119
Db 67 ANGSSSETASSIE-----TCPL-VKEY-----YERVNGDGKRHPCKRDKAKNEDVNRFS 114
QY 120 DESEEG-NKIRNYKRNDAI-ACAPRRHMDCKMLELNDINTQNIHDLGNVLYTK 177
Db 115 TLGGQCTYRKIKSOGDNKVGACAPRRHLCDYNESTID--TSTTKRLLEVCMAAK 172
QY 178 YEGESIVNNH--PHKGT-----SDACTALARSFADIGDIYRGIDMF-----KPNVHDV 224
Db 173 YEGNSI-NTHYTHQRTNEDSASQCLTYLARSFADIGDIYRGIDLVLGIDNKEKEDQRKL 231

```

0Y	225	ETG,REVPFKT,HDGMEDEVND-----YNPDG--SGNYUYL,REAMVNVNRYKVE,ITCDA	278
Db	232	FOKLTADIPFKKH---KQVMTKNGAQBRT IDDAKGQFQOLREDMWMTSNREYWKAL,ICHA	288
0Y	279	SYKSGYFMOSESNT,PLPSNPRCGHKOGKVPNT,NDYVQY,LYRWFDEWGEERFKRRIK,IKLK	338
Db	289	PKEMAY,FKTACNVGKGTNOGCHOCIGSDVPTFYDYVQY,LYRWFDEWGEERFKRRIK,IKLK	348
0Y	339	VKDSGRNDKEKLY,SHNGHDCSTTTIMKKG,ILHLDNKCSTDCSTKCKV,FEVWLCNDOOPAKK	398
Db	349	LOKOC,RDEON,LYCSGNGYDCTKYIKKGLY,IGEHCTNC,SWC,RYATEWIDNOKEFKL	408
0Y	399	OKEKEKEI-----QSYLSDNKR,FVNNINSEY,YKOPEK,LEK,ETQVATN	441
Db	409	OKRKYETREI,SGGSGSGSPKRTKRAANSSSSDNN-----GYSEKTYK,LEK,ETQVATN	460
0Y	442	DTFLNLN,EGKRYCKG---GLPEKIDTFTNSADDKG,IFYRSEY,COVCPD,CGVKCDG,IKY	497
Db	461	DKFLTK,ILNKEG,ICQKQOVGNEKADNVD,FEKUYK--TFSRTEICEB,CWGL,LEK,GPWP	519
0Y	498	THKSDNREKRYNNDY,PRMGVAK,PTNTV,LY,SGNEBGDITQ,LEK,ENFCNSSTYK,KNOK	557
Db	520	KVKGDKT,CGS,AKT,ITYPK---NTD,IP,LY,PKSOON,ILK,RYKNC,CEGAB--GGGQ,IKK	575
0Y	558	MECYK,ND-----ENINBCK,LEONTEI,NDNPK,ILISF,NEFEL,AVTYL,LD,TKYMD,IKLT	612
Db	576	WOCY,DEH,PRSPKNNNN,CGSTGMDK,FGQKQY,KV,SY,NF,FWMDV,HD,LD,SVEMK,TELK	635
0Y	613	CINNTT---HCIDECNR,ICL,CFDRYV,KOKEE,MS,IKK,LT,FKKKN,IOO,SY,SNIN,LF	668
Db	636	CINNTNGT,CTRNNK,CKT,DCG,CFQK,WEK,GOEMMA,ILHD,EGKQD,IVQO-----KGLI	690
0Y	669	EGYEFKYVMD,IKD,KEA,KF,ELME,IKK,KN,EF,NLENNDY,LE-----NAI	714
Db	691	VESP,RYGLDY---LKG,ML,LO,NIK,OV,HD,TD,DIK,IKL,DEBA,VA,VL,GGK,DN,TTI	746
0Y	715	ELLD,HL,KEAT,ICKDNN,NT,ACET,SHNAT,PN,CV,PRG,OT,PK,IK,ELIAOY,FR,SAVE	774
Db	747	DKLLO,HEKEQA,EOCKOR--OEBE-----KKAOQ	773
0Y	775	EARNRGLK,IKL,KG,HA,HEG,IFYR,GG,RRK,DFK,NC,IL,IK,HSN,RL,SG,NS,PCD,CGK,GTG	834
Db	774	ESRGRSAB-----TREDER-----TOOP,ASABE---	797
0Y	835	QTRFVYGT,HEV,ED,HE,RRK,HD,EDY,IM,PR,RR,ICT,SN,LE,HTQD,HP,LC,NG,IV,LD,UNNS	894
Db	798	-----VEEEDDD,DD,DEDD-----DDVQEE	820
0Y	895	FLGDV,LSAK,TEANK,IT,IRAT,KEK,NNL,KG,PEV,TD,PK,HO,TTIC,HA,IRY,SPAD,IG,DI,IRGD	954
Db	821	EEG-----KEB,TV,LE,VE,LE,VE,VEET-----	842
0Y	955	LMERNGD,VK,LO,GLE,TV,FG,NI,HS,KS,KG,ND,RY,ND,AP,RY,LY,LR,EM,WE,AN,RA,RY,WEAM	1014
Db	843	-----	842
0Y	1015	KODI,LYL,KD,KS,GH,OST,SG,SY,CGY,SD,HT,PL,DY,I,PK,IL,MT,EM,EA,MY,CKV,OK,REYD,IKLE	1074
Db	843	-----VTE-----OE	847
0Y	1075	KCKE,CKD,KN,DOG,CT,KS,SG,GT,CT,CF,AC,GE,MY,ND,II,GN,ME,GN,II,SD,XY,KE,HL,HD,QA,MS	1133
Db	848	GVP,PC-----DIY,OKL-----	858
0Y	1135	VSN,SGIE,AS,STAK,NH,ID,RN,VI,EL,SL,ELY,ONG,GS,KN,SGT,SD,ES,AV,IG,NTY,EN,VA,VL	1194
Db	859	-----	858
0Y	1195	HD,TF,FD,CO,OG,SN,FC,DE,CK,SD,GN,DE,KY,AR,DR,PK,OD,HD,AC,CG,KS,SK,TR,PO,IK,TK,KA	1254
Db	859	-----FED,DK,SL,KE-----ACG,IKY,CP-----	875

0Y	1255	EKKTECKTVNDLKENDGKKOYEDCHPPKKNNSGYDQOCGNILNY-----EDPRVCMP	1308
Db	876	-----GGEK-----PNNMCVTPSGVSTATSGKDALICVP	906
0Y	1309	PRROKLCVHFLA-----NDNEIKKIOSQV-----NLKEAFIKSAAEFEFSWYU	1353
Db	907	PRRRRLVYGLSQASRGSGDETEVESSSEATSPASQSESEKLTARFIESAALETFFLMHKY	966
0Y	1354	K-----SKDG-----EGNELDKEIKE-GKTPPAFLASMYTGTGDRDPLFG	1393
Db	967	KEEKKPRATQOGAGLGVSLPEPSPPEBDPTQLOQGVIPRPEFLQMEFTYLAADKDLI--	1024
0Y	1394	TDISKGHEGSEKLEQJDSLFFKNGDQSPGKTRQEMTTEHSHEIMEAMLALVIGAKK	1453
Db	1025	-----	1024
0Y	1454	DDTEPNTGYNNVAFSDKSTLLEAPRPROFLMLTEWYIDCYTROKYKLDVQEKCSND	1513
Db	1025	-----	1024
0Y	1514	OLKODPEONKKCEDYVAKYMKKKKKEMIPDKYKXDEBDKKRPRQHIQVWYDYTGTAETD	1573
Db	1025	-----YSSSNTFS	1032
0Y	1574	YLNKRTASCGDKPGSASVYQRIOLLEKQAVYDADKHCGCTKFTENDKTYTNISSKDKC	1633
Db	1033	-----DTYGKQPPSSNDNL	1047
0Y	1634	KGIKVEANTGAIKQKNGKPNNYNNIKELTFEDVLPFRRLRIGCFHALDGNVDPYKDENG	1633
Db	1048	KNIVLEAS-----	1055
0Y	1694	LKRILMEVAATEGYNLGOYYKEREKEREKIKTSDAHKYSEYEPSCAMKYSFDLIDLIG	1753
Db	1056	-----GSTE-----QEKKKK-----	1066
0Y	1754	IDNLEDEKQKTEENLKTIIFKNKGTSVGKSGSDSTTGPGSTAKRFPNNEKCECVNAMAICG	1813
Db	1067	-----QIOAKIKILN-GATS---GVPVTKNSVTPPOOTWENIANKIWNAMAWCA	1113
0Y	1814	--YKRGDDGNSGNSASADEDLKKCGSVPSDDDYPMGRKNRDEGTAYOFLRAMEWGEDFC	1871
Db	1114	LYTKE-----NDAR-----GTS-----	1125
0Y	1872	KHKEKELEKLVGACNDYTCGDNDENDKRKKCTDACTQYKPFISEMKPOYERQKIKKYEENDK	1931
Db	1126	-----AKIBQND-11333	1133
0Y	1932	IYSEHPAKDAEDAREYLDQKTKKICEKNSGCEYKCMKDYSTORLTDGNSQNMPSASLD	1991
Db	1134	-----LKK-----ALMDANKNTV-----	1147
0Y	1992	EPRVEGKCNCQVPRGPPRVARETPEBPVSLISKATASKREAKTAPTRQPKVEYNTTE	2051
Db	1148	-----	1147
0Y	2052	MRAQTRPRRAAQTRKRTSTATTESDVGMVAKALLSNKPRDSRGIGBCNPKTYGQIPKW	2111
Db	1148	-----	1147
0Y	2112	GCIYVSKSENEGICMPRRKKLCINNIOYLANETENKRDNDIKEAFICAIEFTQFLML	2171
Db	1148	-----IEKYQTVNKKLE-----	1159
0Y	2172	KYIIEHPAENELONGIIPDEPKRIMITYYIGYKXDFPGTGDISNDKKIITYVNSVTTILN	2231
Db	1160	-----D	1160
0Y	2232	ENNNKKODKKDEBELRKIFWEKNKKKFIWEGMIYGLYHLTDENEKIKIDNYQYNDMTKL	2291
Db	1161	ESGAKSND-----TIQ	1171
0Y	2292	TPSLEEPKRPQRLWTFTEMAEEFCNKRKEQLLKLEAGC-----KEYECNGSNDK--	2342

```

Db 1172 PPTLKNEVEIPTFPRWHEWNSFCFERAKRLAQIKHECMDEDEKQY-----SGDGEYCE 1227
QY 2343 -----TOEACAEVYQNFLLKKKTEYEROREKFKDKGKYYKQYPSIERD 2389
Db 1228 EIFSKQYNVLQDSSCAKCRKLYKTIEKKITEYERQKAYEQK-----SNENQOKD 1282
QY 2390 IEKATCAHEVILNMKLKELCGNKDCSCMOKPSOLPKTTOQSSSDANDMESLIDYV-E 2447
Db 1283 -----KC-----QOSNNNAEFSGTLGASPTA 1306
QY 2448 EFNKCEPELSKSGSMHTKKTIPKIPKNCVEKAAAYLSKEANNDITLKEKFIPIES 2507
Db 1307 EF-----LQKLS-----CKNDNGY-----ENGDNKIDRK-NPDKT 1337
QY 2508 TREKESKNSWTNNNCPDKPKPYADKYIGRRNCPENNEENRFKVDYEMKCYKNSKFYQEK 2567
Db 1338 FKEAHS-----COP-CPIGVK-----CONCH----- 1358
QY 2568 KRYCVPRRREHMCRLNIDEKIERLKDSTYLLKMRRTANEGIDILKNENSGCAMP 2627
Db 1359 ---CVGSANGKECKNN--KITAEDIKN-----KTDPNGNIEMVSDSTN----- 1398
QY 2628 ICDPMKYSFADLDIYRGTMRLIGGLPVEIKLYKPEYIYGKRNKRNKGNKYNIDVO 2687
Db 1399 -----TFEHLGD-----CKSSGIFKGI----- 1415
QY 2688 TFRSAMWANDNRKDIWK-AMTCKAPBEDAKLFRKGRMDGFERITLIOD-KCGHKDDPPVDY 2745
Db 1416 -----RKDEMKCANVC-----GVDICILEKKIKNGQBED---KIT 1447
QY 2746 IPORFWMTEMSRYCYKALMEELERKKSQDHC-KTSDRKNDYDENKCEQCTRCQERYK 2804
Db 1448 ITWK-ELKRWLEYF-----LEDYNRIRIKRIKICTKKEDECK-----CIKGCEI- 1490
QY 2805 NEVLAKWSLFDIOSNKRYKELEYEQITYTKISTYDHOVNFQOKLTPEKSECSVSFSYLYE 2864
Db 1491 ---KW-----VQ-----EKREMO----- 1501
QY 2865 TSKCLNKKENENDSSNITYAEETPKSYKEACSTLPSKNPLDNCPTDNKDGCKELQ 2924
Db 1502 -----KIND-----TY-----LE 1509
QY 2925 TFFPCKNDYDNNDNMNAYLYLNSDDNKGVLIIPRRRHLCRPTTAYNRKDGKEILK 2984
Db 1510 QY---KNDDGNTLNF-----LEQOYR----- 1529
QY 2985 KLLITSASOGULLQOKYKSEELCEFAKMSYADYSDIITKGTMDMDTSLSEKIKIFET 3044
Db 1530 -----TEFKNAIKPCDGLD----- 1543
QY 3045 SNEATENRKTWENNRRQIWMHMLGKYKATSKYVLDEGWQLPDEETNOFLRWLIEWA 3104
Db 1544 -----OF----- 1545
QY 3105 KOACEKKNVSDSLTKGPRSNEDNFEASELLRQPCQNDIKYISLNLIKNTMENLNI 3164
Db 1546 -----KTSGLINSTDN-----SQNGNNND-----LYCLINKLQK 1575
QY 3165 KYKOLKODSSGINDKPSREBNVQYIKSDSCALELNDIMEITYGTNNENN----- 3217
Db 1576 KISEKEKHSQGTQF-PCD-----NSLSCKESYLVEDVDYEE-----ONPKKVVQPKFC 1626
QY 3218 -EFKVLKLLPGLYFVDETHKNHVLGNT--KEEDQTVPRKALYFTPHVDSYQAPL 3274
Db 1627 PDMKPKKENDVEVGTCCGDEKKKVEDSVLEQKEEFAASAPESPPLTP-----EAPK 1680
QY 3275 FSTRHVAQYD-----KN-----DILKSSISVIVLSAGL-----IALHFKMKKFKFS 3316
Db 1681 KEENVVPRPPPKRRIRIKTRNVLDHPAVIPALMSSTIMWSIGIGFAFYFLAKKTKKS 1740
QY 3317 SV-DLRLTNLPOGEYGMPTLESKNRYIPYSGPKKGTIYIMGDSGDEKYMMDLSS 3375

```

```

Db 1741 SVGNLFQIILQIPKSDYDIPFLKSNRYIPYADRHKGTIYIEGDSGDE-KYAEWSDT 1799
QY 3376 SDITSSSESEYEELDINDIYVPGSPKRYKTLIEVLEPSKRDIUSD-----DTPSNDFPR 3428
Db 1800 TDIITSESEYEELDINDIYVPGSPKRYKTLIEVLEPSKRDIQNDINDIIPSDIPNDTTP 1859
QY 3429 TNRFIDENNELKHDVSVQYL---PTEPRNNNTKSDIDIMNTEPNTLYSDNPEKPFITIS 3485
Db 1860 P--ITDDEMOKLKDPIISNMLQNTQTEPNILHDVND--NTHP-TMSHHNDQKPFITIS 1914
QY 3486 IHDRDLYTGKREISY-----NINMSTN-----TNNDIPMARNDSTYRGIDLINDS 3529
Db 1915 IHDRNLFSGEYNYDMEFNGSNPNINISDSTNGWDSLTLSNNHSPYDKNDLYSGIDLINDA 1974
QY 3530 LVLNLTLY--MAK 3541
Db 1975 LSGNHIDYDEMILK 1988

RESULT 11
AAM93944
ID AAM93944 standard; protein; 2228 AA.
AAM93944;
30-JUN-1999 (first entry)
P. falciparum PfEMP1 protein.
Erythrocyte membrane protein; EMP; PfEMP1; malaria; antioocclusion;
glycosamino-glycan-like moiety; antiaggregational; antimalarial;
antigen receptor; infected erythrocyte; rosette formation; blood cell;
capillary occlusion; cerebral malaria; treatment; vaccine; detection;
medicament; parasite; diagnosis; drug screening.
Plasmodium falciparum.
W09915557-A1.
01-APR-1999.
18-SEP-1998; 98WO-SE01675.
19-SEP-1997; 97SE-0003386.
(KARO-) KAROLINSKA INNOVATIONS AB.
Barragan A, Carlson J, Fernandez V, Qijun C, Wahlgren M;
WPI; 1999-254692/21.
New isolated malaria polypeptides
Claim 4; Page 67-74; 80pp; English.
This invention describes a novel Plasmodium falciparum erythrocyte
membrane protein (EMP), PfEMP1, which is capable of binding to a
carbohydrate which exhibits at least one negatively charged
glycosamino-glycan (GAG)-like moiety and has antiaggregational,
antioocclusion and antimalarial activity. The carbohydrates of the
invention are capable of acting as receptors for malaria antigens
present on the surfaces of malaria infected erythrocytes, by binding
to these antigens the carbohydrates prevent erythrocyte formation by the
blood cells, this prevents occlusion of capillaries as is seen in
cerebral malaria caused by Plasmodium falciparum. The products of the
invention can be used to treat malaria or to vaccinate against it, or
used to design a model to identify compounds that bind to PfEMP1. The
carbohydrates, polypeptides and antibodies of the invention can be used
as a medicament for dissolving the rosettes formed by erythrocytes
infected by a malaria parasite. The products can also be used for
detection, diagnosis and drug screening.
Sequence 2228 AA;

```


Db	1667	GTCTGCKEKKDE-----	-KK--EESE-----	1668
Qy	3055	WVENNRQIMHAMLCGYKIATSKVTLDEGWCQLPKDEETNOFLRMILIMAKOACEKKHV		3114
Db	1687	-----	-----PAEESE-----	1702
Qy	3115	SDSLTKCPRSNEDNEFASSELLRPGCGQNDIRKYSILNLIKNTMENLNKKYKQKQDSS		3174
Db	1703	AASEET-----ETNFP-----DEPG-----		1717
Qy	3175	GNIDKRPSEENVOSTIKSKSQCALELNDINEIYTGTKNNENNEKEVILKLYPLGYFE		3234
Db	1718	-----	-----TGPAAPSTPAPPT-----	1731
Qy	3235	DETHKNHYLDGNIKEEQTVRPKALYFETPHVDSFYQAPLFTSHAVQYDPKNDILKSI		3294
Db	1732	-----PDPPTPLRPOA-----DEFPDST-----ILQTTI		1755
Qy	3295	SVYVYSLALGLIALHFMKKRKSIV-DLRLTIPOGEYGMPTLESKNRIYPRSGPYKKG		3353
Db	1756	PGVALALGIAFELFKKTKKASVGNLFQILQIPKSDYDIPFLKSSNRYIPVSDRYKKG		1815
Qy	3354	TYIMEGDTSGDECKYKWMWDLSSDINSSSEKVELDINDIYVPGSPKKTILEVYLEPSK		3413
Db	1816	TYIMEGDS--DEDKTAFMSDITDVYSSEYEELDINDIYVPGSPKKTILEVYLEPSG		1873
Qy	3414	RDIPSD--DTPS-----NDTPRTNRFTIDDEMNELKHDVFSQYLPNTEPN--NNKSA		3461
Db	1874	NNTYASGKNTPSPDRIQNDIGPSKRTIDNEMNQKKRFEISMNLQN-QPDVYNDNTTSG		1932
Qy	3462	DIPNTE-----PNLYSDNPEEKPFIIISHNDLYTGKEISYINAKSTNTN		3508
Db	1933	NSSTNTNTTTSRRHNVNDNTNTMTSMDSNMEENLLPSIHGNTLYSGEYSYVNM-VNSM		1991
Qy	3509	NDIPMANRNDYRGIDLINDSL		3530
Db	1992	NDPIPNEDNNVYSGIDLINDSL		2013
RESULT 12				
AAW00385				
ID AAW00385 standard; Protein; 1726 AA.				
XX AAW00385;				
XX	21-FEB-1997	(first entry)		
XX	Truncated Plasmodium falciparum erythrocyte membrane protein.			
XX	Plasmodium falciparum; erythrocyte membrane protein; malaria; detection; identification; treatment; prevention; parasite.			
XX	Plasmodium falciparum MC type.			
OS	Key	Location/Qualifiers		
XX	Domain	62..394		
FT	Region	/label= Duffy binding ligand domain 1		
FT	Domain	607..648		
FT	Region	/note= "Cysteine rich motif"		
FT	Region	/label= Duffy binding ligand domain 2		
FT	Region	1488..1523		
FT	Region	/note= "Cysteine rich motif"		
XX	MO9633736-A1.			
XX	31-OCT-1996.			
XX	26-APR-1996;	96MO-US05798.		
XX	27-APR-1995;	95US-0430908.		
XX	(AFYV-) AFYVMAX TECHNOLOGIES NV.			

XX Baruch DI, Howard RJ, Pasloske BL;
 PI WPI: 1996-497376/49.
 DR N-PSDB: AAT41853.
 XX
 PT New Plasmodium falciparum erythrocyte membrane proteins - used to
 PT develop products for the diagnosis, treatment or prevention of
 PT malaria parasite infections
 PS Claim 1; Figure 12; 149pp; English.
 XX
 CC A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte
 CC membrane protein 1 (PfEMP1) or active fragments or analogues of that
 CC protein can be used in the treatment or prevention of symptoms of a
 CC malaria parasite infection. The polypeptides can inhibit, block or
 CC reverse the sequestration of erythrocytes in patients suffering from
 CC malaria. Nucleic acids derived from the PfEMP1 gene can be used as
 CC probes and primers to identify a Plasmodium falciparum parasite, the
 CC primers used to generate characteristic amplification patterns from
 CC immunoreactive with the PfEMP1 polypeptide or its fragments may be
 CC used in diagnosis of malaria infection. This is a truncated PfEMP1
 CC protein of the MC type of Plasmodium falciparum. The full length
 CC PfEMP1 protein is given in AA00384.
 XX
 SQ Sequence 1726 AA;
 Query Match 6.4%; Score 1244.5; DB 17; Length 1726;
 Best local similarity 22.2%; Pred. No. 1.4e-65;
 Matches 501; Conservative 276; Mismatches 632; Indels 843; Gaps 88;

QY 33 KSARNVLERYAKNI-----RHPSYAKEHVDLSKGLDTRAEFRGSPSPVKNHNYPPY 86
 DB 13 KDAKHALDRIGEYVYKEKYENDAEKYYK-----ALKGNLQAKGIGELASS-----PN 60
 QY 87 PCMLDKRHEHNLAHYDDVNLRHCHGREGONRDEDESECC-NKIRYKKNDAI-ACAP 144
 DB 61 PCFL-VEYYNNRLK--RRKYPCANROYVSEYGGQCTFNRIKSENNDSLGACAPY 117
 QY 145 RRRHMDKNLEALNDINTONIHDLGNVLTAKEGESIYNNH-PHKG-----SDACTA 198
 DB 118 RRLHLCDYLNLEKMGKSTYK-HGLLDVCAAAEYEGDSITHTTKHELTNPDRKSQCTI 176
 QY 199 LARSFADIGIVRGIDME-----KPNVHDVETGLREVEFKIHGMEDEVKNDYDPGS 252
 DB 177 LARSFADIGIVRGIDLYDDEKDERKLENNLEIFKLIHENLGTQADADHYKDE 236
 QY 253 GNTYKLEAMNNVNRKRYWEATTCDSYKSGYMOSESNTPLSNP-KCHGKGG----- 306
 DB 237 ENYYQLEHEDMTANRSTVWAITCHGESDPRFKTCCSGEWTDDKCRCKDEGKMETNE 296
 QY 307 VPTNLVPOYLRFMDEMGEEFCRKNRIKLKVKYKDSGRNDK--ERYLCSHNGHDCITIM 364
 DB 297 VPTVFDVPOYLWFEEMADDFCRKRRKKTLENAIKNCRGKGNDR-YCDLNGINCEETAR 355
 QY 365 KRGILHLDNKCTDSTCKKFEYVLGNQDEAFKKOKKEKYEKIO-----SYLSDNKEVNN 420
 DB 356 GAELFVKGDCCHGCSVACDFVAMIDNQREKFEKKYKDEINKTGTITTINGNINN 415
 QY 421 INSEYKQFKEKLEKETOYATNDFNLNLNGKYYK--GGPGEADITFTSADKGIFFYR 478
 DB 416 L--YVGHFELDKR-YYPVDSLOKLNDEALCKPPNNGNEKASVVDNNEVNTFSH 471
 QY 479 SEYCOVCPDGVKCD-----GIKTYHKS-DNDRREVNNEDYKPPAGVYPTNITVLYSNGEO 533
 DB 472 TTYTCEAPMGCAQKKEKNGGKAKKESCAKKERIFPK-----NSTDIKILTPKGR 524
 QY 534 GDITQKLENCNSSTNYKKNKNOKECYKNDENINRCKLEPONTIEND-----N 582
 DB 525 SKTELEKLTCKDQOKIK--NDIMKCHYDNG-----TDDQTDSDSDCYLGDWGNLTKE 576
 QY 583 PKIISFNFELWYTIILRTIKMNDKLTCTIN-NTTTHCIDECNRNCLCFDRWVKQKEE 641

DB 577 DKIMSTNAEFWMVMDHMLDLSIKWRDEHGRCLIKDKGKTKICGCKNKKJCIFQWVQDKPT 636
 QY 642 EWNISIKKLEFKKKNKIOOSYNSINNLFEYGFYKMDIKRDEAKKRELMENIKRKNRES 701
 DB 637 EMGKIDHFRRKOKDIPKDW----- 655
 QY 702 NLENNDYLENAI--ELLDLHEKATATICKDNNTNEACETSHNATNPVCKPRGTOPTK 759
 DB 656 ---THDFLOTLTKMLDLEIITDY-----GDANE----- 683
 QY 760 NIKELIQVFRSAVEEARNGHLKLGKAEGLYKKGGRKDKDKNLCRIMIKHSNRNLG 819
 DB 684 -IKRI-----EALLEQA--GVGGIDFPAALAGLYTRGFVAE--KDTTDLKLQHEOK--- 729
 QY 820 FSNGPCDKSGTGDGIOTRFVYVGTMEVDEPRHMRKHEDVIMPPRRRHICTSNLEHLQTD 879
 DB 730 -----EADKCLKTHTDVCPPO-----ED 748
 QY 880 HPLNGNIYDVLNNSPLGVDLLSAKYEANKIIMYKEKNLKGPKVEYDPRKHOTTCRAI 939
 DB 749 RSV-----ARSESATV-----PSPPADRK----- 767
 QY 940 RYSPADIGIIRGRDLMERNRGMVKLGHLLEYVFGNHLKSGKGNKYNDAPKYLKLR 999
 DB 768 ----- 767
 QY 1000 ENMWEANRAKWEAMKCDIKYLKDKSGHSTQSSYSGSDHTPLDYIPQKLRMTEMAE 1059
 DB 768 ----- 767
 QY 1060 WYCKVOQKEYDKLEKCKECKDKDNGOGCTYKESGTCTCTEACNEYNDIILGKKDOWNI 1119
 DB 768 ----- 767
 QY 1120 ISDKYKELHQAOAMSVNSGSIKASTAKNHDNVIEFLSELVQOANGKNSGTSDESA 1179
 DB 768 -----ATEVDAN-----ASSDE- 781
 QY 1180 VIGNTTYENVGAYLHDGTGNFDCQSQNEFCDEKSDKDNKRYAFRDKPODHGAGCGKS 1239
 DB 782 -----DDEFEEREE--EEDDEEEAEVQOEKTP--ESATPAVA 816
 QY 1240 GSKPTRVQIYTKKKAEEKDTE-CKTVNDLIKENDGKQVDECHPKKNSNGYPMQCGNIN 1298
 DB 817 PSPPGTQDQVKKPASQEDVYKCSIYDKALK--GKLD-DACILKYGTAPTSMKC--- 868
 QY 1299 LVEDPR-----VCMPPRRQKCYHFL-----A 1320
 DB 869 IPSDTKSVATGSDTGTGSGSICVPPRRKRLVGLKLDHMACGETTEAKSQETSGGQKTPSG 928
 QY 1321 NDN-----EIKKIQSOVNLKEAFIKSAAEFPFSWYKSK----- 1356
 DB 929 NESPSPSKLPQGPRTPTETTKTPESSLSLHAFVSPPRLRPLPWHKFKQMAQAQACATGL 988
 QY 1357 -----DGEENELDELEKGIIPAPFLRSMFYTFGDRFLFTGD-----ISKG 1399
 DB 989 QLPGVYVDSDDPQOQLKRGNIIPNDFLRQMFYTLGDRDICIGGDRDIVGDTIVSITTEG 1048
 QY 1400 HGEGRKLEKQIDSLFKKGGDKSPNGK-----TROVMWTEHSHETIYEA 1442
 DB 1049 ESTRKIKSLIEGFLKQVYSPSPROTSSRTVPVHPQTSVEKTPQQTWMBANGPHIWMG 1108
 QY 1443 LCAALV-----KI-----GAK-----KDDTENYGVNNVFKSDKS 1471
 DB 1109 ICALTYDSGAIGOPQKVEDADKYLEKLKPNANGIKWMLKEDNTSSAMPTSSSSSGS 1168
 QY 1472 T-----TLSEFAKROFLMLTWMYDYCYTRYOKYLKDOVERCK----- 1510
 DB 1169 NDIPTNPLTEFEVFIPTFFRYLHEWGCNFCERKRLKQIYKECKVGENGYGRGRQKTP 1228
 QY 1511 -----SNDQLK-----CDTE--CNKCEBYVYMKKKK-EMTIPQK-----Y 1545

Db 1229 QCSYGEDCEDLSKSYDYVADIECPKCAHCWYKMYIKKKDEFEQKAPKQKDY 1288
 QY 1546 KDERDKRRDR-----QHGVAVDYDTGTNATDY-----NKKFTASCG 1584
 Db 1289 VGNNGKGGNGFCITLKSLSDAACFLERKISCKKDNSENGNDKINFSPNETFPATY 1348
 QY 1585 DRGSAVYQNRNIQLEKQAYADRHCGCTKFIENDDKYTNSSDKCKGLYKANTGA 1644
 Db 1349 CRKCSFKIDCKENGCKCKNGGGGTNETCGTFTY-----TSENFKQNG-QTAKERY 1398
 QY 1645 IKWQNGPNYNNLKELETEDV-LFESRL-----RICEHAL-----DGNATDPYVKDE 1691
 Db 1399 MRVSDNNPNGFDLNEACQAGIFKSIKRDMECGVCCGYCIVIPKNGCVTTSGEANDQ 1458
 QY 1692 -----NGLRRIMEVATTEGNTAGQYKKEKKEKITSAN-KSYEVPPC----- 1737
 Db 1459 IITIRLVAHWQ-NFLDYD-----KIKHKISHCKNSSEGYTCIKNVEQWI 1505
 QY 1738 SAMKSYFDLRDIILIGIDNLEDEKTEENIKKI-----FNKN-----GTSVAKGSD 1784
 Db 1506 STKRTWTNIK-ILN-----EYKNDPNYNVKTILLDLOSQIDFNKAIRPCGT-LTKFED 1559
 QY 1785 STTGNGSTARRFENNEKCYWNNAMICYKGR--DDGNSGNSARSDDLKCC--GSY 1839
 Db 1560 SSGINGAESSEKKNH-----YDAIDCMNLRLQDKIDCKNNHAGNGENAKCKEHA 1614
 QY 1840 PDDDYPRKKNDEGTAYOFLNFAWGEDFC-KHKEKELEKLVACNDYTCGDNEDKR 1898
 Db 1615 PDEDDDEALEENPYQ-----PNICPKPEPRAEKKG-----CEPAEKKEK 1657
 QY 1899 KCTDACTQYKFKISEWKPOYERKQIKKYGKNDKITYSEHPV-AKDAEDAREYLDQKKI 1956
 Db 1658 V-----EKEEKTVMNTVAKPTEKEAAGDAGAAADSEENP----- 1693
 QY 1957 CENKSGDCYKCKMKDYSTORLTGDSQNPAS 1988
 Db 1694 -EKAPEPEVETKKDKAPVKT-----PAS 1717

RESULT 13
 AAB18144
 ID AAB18144 standard; Protein; 1700 AA.
 AC AAB18144;
 XX 07-NOV-2000 (first entry)
 Dt
 XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:1.
 DE
 Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine; antimalarial; malaria; protozoacide; infection; insecticide.
 XX Plasmodium falciparum.
 OS
 XX WO200025728-A2.
 XX 11-MAY-2000.
 PD
 XX 05-NOV-1999; 99WO-US26796.
 PE
 XX 05-NOV-1998; 98US-0107131.
 PR
 XX (HOFF/) HOFFMAN S.
 PA (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 XX
 P1 Hoffman S, Carucci D, Gardner M, Venter JC;
 XX WPI; 2000-365347/31.
 XX
 PT Proteins encoded by chromosome 2 of the human malarial parasite, Plasmodium falciparum, useful as antimalarial vaccines and in the

PT diagnosis of P.falciparum infection -
 XX
 PS Disclosure; Page 29-33; 577pp; English.
 XX
 CC The present invention describes proteins and their fragments (I) encoded by chromosome 2 of the human malarial parasite, Plasmodium falciparum. Also described are: (1) nucleotide sequences (II) encoding (I); and (2) vaccines against P. falciparum infection comprising (I) or (II).
 CC (I) and (II) are useful for the development of vaccines against P. falciparum infection. (I) and polyclonal antisera or a monoclonal antibody raised to immunogens comprising the sequences of (I), are useful in the detection of infection with P. falciparum. Furthermore, (I) (especially when they are rifins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent P. falciparum infection, or they can be used to identify drug resistance in P. falciparum. Sequencing of the Plasmodium chromosome 2 and the subsequent identification of proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the complexity of the parasite life cycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new drugs. AAB70078 to AAB70287 and AAB18144 to AAB18352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification.

Sequence 1700 AA;

Query Match 5.4%; Score 1045.5; DB 21; Length 1700;
 Best Local Similarity 16.0%; Pred. No. 1.3e-53;
 Matches 576; Conservative 198; Mismatches 622; Indels 2195; Gaps 83;

QY 33 KSARNVLERYAKNI-RHPSKYAKEHVDLSKDLTFAEFRGGPSPVKNHNTYPP----- 85
 Db 13 ESKVDFLRIGKRYEKEKIAKRTTELHGLDSATY-----PNDKH-----FBGSTEN 62
 QY 86 YPCNDLHEHNRILKRDVYLNKPCRGKQNRNDEDESECGNKIRYKKNDAIACAPR 145
 Db 63 NPKQLQYDNTNVTN-GEQGEYPCETDIYERFSPDEGAQCDK--KIKDSEGAAPYR 118
 QY 146 RRMCDKMLEALNDIN-QONIHDLGNVLYAKTEGESIVANNHP-HKGT-----SDCTA 198
 Db 119 RLHVYVRNLENTINDYSKNNKHNLLVEYCLAKEGESITRTYPOHDETNPDKSQLCTV 178
 QY 199 LARSFADIGDIYRGIDMK-----PNVHDEVETGLREVEFKIHDM-----EDEYKN 245
 Db 179 LARSFADIGDIIRGMDLYRGNTKEKKKKKLEMLKTIPIHITIDELKNGTNEEELQK 238
 QY 246 DYNPDGSGNYKKLBAWNNVNRKWEAITCDA-SYSGTYPMOSESNTPLFSNPKCGHKQ 304
 Db 239 RYRGDKNDYFLKREDWWDANRETYWKAITCMASGYOSOPTCGRGEIPYVTLKSCOCIA 298
 QY 305 GKVPNTLDYVQYLRWFEDWEEFCRRKNRIKLKVKDSCR--NDKERTVCSHGHCTT 361
 Db 299 GEVPTFYEDYVQYLRWFEMAEDEFCRKKKKIIPVKNTRCROVGRKER-YCDRGGVNCDG 357
 QY 362 TIRKGIHLNDKCTDCTCKVEFWLGMQEAFFKQKEKEKEIOSYLSNDKFFVNNI 421
 Db 358 TIRKQYIYRLDPTCTCKSLACTFAEWINDQKEQFDKOKQYQNEISGGGGRQRKSTHS 417
 QY 422 NSE---YKQFYELKLEQYATND-----TFNLNLNGSKYCKGLP-GEEDITFTNSADD 472
 Db 418 TKEYEGTEKHEFNEELR-----NEKGRVRSFQLLSKEKICKERIOVEETAYNGENE 471
 QY 473 KGIFYREYQVCPDGVCKDGIKYTHKSDNDRVRANNEDYKPPWGVKPTNITVLSGNE 522
 Db 472 SNFESHTEYCDRCPGLGVDC-----SSDN----- 495
 QY 533 QGDITQKLENFCONSINYKDKNNQKWCYKKNENINCKLEQNTETINNDPKIISPHNF 592
 Db 496 -----CPFB----- 499
 QY 593 ELAWTYILRLTIKWNDLKTKCINNTT-----THCIDBCNKNCLCFDRWVYQKEEENNSIK 647

Db 500 - - - - -GCGCAGCTGTAGTGGCGCATAGTTTAC - - - - - 531
Qy 648 KLETKKNNIQOYSYNNINLPEGTFEKKVMDKLDKDEAKKELMENTIKRKNESNLNNR 707
Db 532 - - - - - 531
Qy 708 DYLENAIELLDHLKETATICKDNNTNEACETSHNATTPNCVPRGTOPTKNIKEIAQY 767
Db 532 - - - - -GCGATGAAAGTGTCAAGATTAT - - - - - 555
Qy 768 FKRSAAEAAARRGLHKLKGAHGIYKRGRRKDFKDLICRIMIKHSNRMIGFSGNCPDG 827
Db 556 - - - - -TTGATAG - - - - - 562
Qy 828 KGTGDSIQTRFVVTGEMVEDPEHRRDHEDVIMPRRRHICTSNLEHLQTDHPLNGNIY 887
Db 563 AATAGGGAAGAAAGT - - - - - 577
Qy 888 DDLYNNSLFDVLSAKYEANKIIRMYKEKNLKGREVTDPKHQTTICHAIRKXSPADIG 947
Db 578 - - - - -TTACGAAAAAAC - - - - - 589
Qy 948 DIIRGHDLMERNGDVYKLGHELFEGNIHSLKGNDRKNDAPRYLKLRENNWEANR 1007
Db 590 - - - - -AGA 592
Qy 1008 AKYWEAMKCDIKYLKDKSGHSTOSSYCGSDHTPLDDYIPOKLRMTEMAEWYCKYQK 1067
Db 593 AAAGATTGC - - - - -AAAAAGAT - - - - - 609
Qy 1068 EYDKLEKKECKD - - - - -KONGGCTKRESGT - - - - -GCTGK-TEACNEYNDIIGLMKE 1115
Db 610 - - - - -ATACTACTGAATTCGATGTGAT-TTGCAAAAAGACATATTC - - - - - 651
Qy 1116 OMNIISDKYELHEOAMSVNSGIEASSTAKNHDNRNVEFLSELYOONGSKSNKSGTS 1175
Db 652 - - - - -CAATGATAAACAT - - - - -CCGAAGAGATCA 678
Qy 1176 DESAVIGTNTTYENVAYLHDITGNFDCCOSQNEFCDEKSKGNENKAYAFDKPOHDGAC 1235
Db 679 CACAAATATATCCATG - - - - -CAAACTTCATATGATTAATTAATGATGTTAC 727
Qy 1236 GCKSGKPTVVOJTKKKKAEEKDECKTVNDILKENDGKQOVDECHKKNKSNVPMQCG 1295
Db 728 TCAATGGTTT - - - - - 737
Qy 1296 NINLVEDPRVCMPPROKLCVHFLANDNEIKLQOVNLKEAFIKSAAETPFSWYYS 1355
Db 738 - - - - - 737
Qy 1356 KDEGNELDKEKEGKIPPAFLASMTTFGDIYDFLEFGTDISKGHSGSKLKQIDSLFK 1415
Db 738 - - - - -GCTCAAGAGAT - - - - - 749
Qy 1416 NGQKSPNGKTRQEMWTEHSHEIWEAMLALVIGAKKDFTEYGVNNKFPDQKSTTLE 1475
Db 750 - - - - - 749
Qy 1476 EFAKROFLMLTEWYDYCYTRQKYLKQOEKCSNDLCTECNKKCEDYVYKMKK 1535
Db 750 - - - - -CCT - - - - - 752
Qy 1536 KEMIPQDKYKDERDKRRFRQIHGVWVITYTGTNATDYINRKFYASCGKPSGSASVYOR 1595
Db 753 - - - - -TGGA - - - - -AACGACATAGT - - - - - 769
Qy 1596 NIOLLEKQAYYADRHGCTKFIENDDKYTNISKDKGLVKEAANTGALKMOKKGNNY 1655
Db 770 - - - - -AGACGTTT - - - - -TTGATACAGAGAG - - - - - 795
Qy 1656 NMLKEITEDVLFPSSRLILCFHALDGNVTDPEVKDENGKLRLEAVATGEGYNIGQYKE 1715

Db 796 - - - - -CACATGTGATA - - - - -GAAAAAATAAAGATA - - - - - 826
Qy 1716 KKEKEKITSDAKHYTEVPCCSAMKISFYDLNDIILGIDNLEDEKOTEENLKIFPNK 1775
Db 827 - - - - -TA 828
Qy 1776 GTSVKGSDSTGCPGSTARKEFPWENKECVMNAMIQYKRGNDGNSGARSDEDLK 1835
Db 829 GTGAGAGAGCTTG - - - - -CG - - - - -CTC 846
Qy 1836 CGSVPSDDDYPMKRNDEGTAYOFLRWFAMGDEFCHEKELEKLYGACNDYTCGDNE 1895
Db 847 CATATA - - - - -GAGC - - - - - 856
Qy 1896 KKKCTDACTQYKFISEMPQYEQIKKYENKDIYSEHPVAKDAEDAREYLDQKK 1955
Db 857 - - - - -ATTACA - - - - - 862
Qy 1956 ICENKSGDCEYKCMKQVSTORLTDGNSQNNPASIDDEPREBKCNQVPRGPRVRET 2015
Db 863 - - - - -TGTAT-GCGTAGAAATTGG - - - - - 882
Qy 2016 PSPRVSLSKATASKKEAKTAPPTKOPKVENLTTEMRQOTRBRRAOQTRKRTATTT 2075
Db 883 - - - - -AAATATCAATGATTA - - - - -TAGTAAATTAATAATAAACATATAT 924
Qy 2076 ESDVGTWKAILSNKPSRGIEGCPNKTGYQYPKWGCIVGSKENENGICMPRRKKLC 2135
Db 925 TATTTGAGAGAGTGCTGCTTGACAGC - - - - -CAATATGAAG - - - - - 961
Qy 2136 INNIOYINTEKRRDDIMEAFIKCAIETQFLMLKLIENPAEENLQNTIPDEFK 2195
Db 962 - - - - -GGAATCAATA - - - - -CAGTC - - - - - 978
Qy 2196 IMYTYGDKMEFGTDISNDKRIITVNSVTTLNENNRKODKKDEELRKIFWEK 2255
Db 979 - - - - -GT - - - - -TKTDROQE - - - - -FMGTG 995
Qy 2256 KFIWECHITGLYHLIDENKEKIRDNYYNDMT - - - - -KL-PSLEEYKROFLRMTEW 2311
Db 996 KDIWKMLCA - - - - -LOEAGGKTLTETYSNVSFNGHLGTKLNEBASPSFLRMTEW 1051
Qy 2312 AEEFCNRKROLLKLEAGCEYEBGNG--SNDGKTQOEAECVYQONIKMKTEYEQRE 2369
Db 1052 GDQCFREIRITQOLILKRCWYQYQNGDKKODKKECTEACTYKEMLTWMDQYKQNO 1111
Qy 2370 KFKKDKKKYKDYPTSTERDIEKATCAHEYLNMKLE-LCGND--CSCMOKPSSQLP 2424
Db 1112 RYEVKGTSPYKE--DSQVKEKYAHGYLRKILKIMICTSGDIAYCAME-- 1160
Qy 2425 KTIQOOSDAMPESLDYVPEEFNK-CQPELSKSGSMITHKITEKIPMNCVEKAA 2483
Db 1161 -GSTJDSNNNDNIPESLKYPLIEIEGCTCKPDS-GEVYIPEKKYVEPV-- 1209
Qy 2484 YILSKEAENMDITLKKFPIESTKEKSNSTNNPPDKPRVAPDYIGHRNCPEN 2543
Db 1210 - - - - - 1209
Qy 2544 REENRFKVDYEMKCYKNSKEYOEKRRVCVPRRHEMCLRLNDEIKIERLDSNYLKMVR 2603
Db 1210 - - - - - 1209
Qy 2604 RTANEGIDILKFNSENGCANPICDTMKYSFADLDIYRGDMLRIGGLPVEIKLY 2663
Db 1210 - - - - - 1209
Qy 2664 KVEFYIYGKRNKKGNKYNVQYTPRSAMWANDRNDIKWAMTKAPADAKLFRKGMGD 2723
Db 1210 - - - - - 1209
Qy 2724 FERITLIQDKGHRDPPVDYIPQRERMTEWSEYCAKLMELKFKKSDCHKTSR 2783
Db 1210 - - - - - 1209

```

OY 2784 GANDYENKCEQCKTCQGEYKMFVLMKSLFDIQSNKYKELYEQPIYKISTYDHVQNFV 2843
DB 1210 ----- 1209
OY 2844 OKLTKFKSCSVSESEYELHETSKLNTYFNENDSSNIRTYAFETPKSYKACGCTLP 2903
DB 1210 ----- 1211
OY 2904 SKNPNDNCPDQNDCKCKELQTFETCSKNDYNNLDNNMAYLVNSSDDNKGVLIPRRR 2963
DB 1212 K----- 1215
OY 2964 HLCSTRITAYNRKGDKEILKLLTSAPSGQLGQTKSEELCEFAKMYSVADYSOI 3023
DB 1216 ----- 1215
OY 3024 IKGTDMMDTSLSEKIKIPETSENEATENRKTWMENNRQIMHAMLGKYKATSKYTLDEG 3083
DB 1216 ----- 1215
OY 3084 WCQLPREDETNOFLWMLIEMAKCKEKKHVSLSKTCPRSNEDNFEASELLRQPGCQN 3143
DB 1216 ---LPK----- 1221
OY 3144 DIRKTIISLILIKNTMENLNIYKOLKQDSSGNIDKPSSENVQSYIKSKDQCALEND 3203
DB 1222 ----- 1225
OY 3204 INEIVTGKNNENNEFEVYLKLPGLYFVEDETHKNHVDGNINKEDEQTVAPKALYFT 3263
DB 1226 ----- 1228
OY 3264 PHVDSFYQAPLESTHRAVQYDPRKNDILKSSISVIVSALGLIALHFMKKKFSVDLARI 3323
DB 1229 PAL----- 1269
OY 3324 INIPGEGMPTLESKNRYIIPRSGPYGKTYIMEGDT--SGEDKYMIDSSDIT- 3379
DB 1270 INIPSSDIDIPKLPSPNRIIPTSGYKRGKRYILEGSGDUSGYTDHY-----SDTTS 1323
OY 3380 SSESSEYELINDIYVPGSPKYLIEVLEPSKRDIPS--DDTPS-----NDTPRPN 3430
DB 1324 SSESSEYELINDIYAPAPAKKTLIEVLEPSGNNITASGNNTSPDQNDIQNDGIRSS 1383
OY 3431 RPIDDEWELKHDEYSQYLPNTEN---NNKYSADIPMTEPNTLYSDNPEEKPIISIH 3487
DB 1384 KITDENMTLTKDEFISQYLOSEPMDVPNDYSSGDIPLNTOPNTLYFDNPDEKPIITSIH 1443
3488 DRDLYTGKEISYNINMSTNTNDIPMANRNDYRGIDLINDSLVYVNLIIY 3538
DB 1444 DRDLISGEYSINVM-VNTNNDIPISGKNGTSGIDLINDSLNSNNVDIY 1493

```

```

PF 07-SEP-1994; 94WO-US10230.
XX
XX 10-SEP-1993; 93US-0119677.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Chitluis C, Miller LH, Peterson DS, Sim KL, Su X;
XX Wellens TE;
XX WPI: 1995-123427/16.
XX N-PSDB; AA083528.
XX
XX New erythrocyte binding domain polypeptide(s) - isolated from
XX Plasmodium binding proteins, used in diagnosis, treatment and
XX prevention of malaria
XX
XX Disclosure: Page 56-57; 81pp; English.
XX
XX Erythrocyte binding ligand (EBL) family genes were cloned from
XX P. falciparum chromosome 7 subsegment libraries constructed during
XX genetic studies of the chloroquine resistance locus. The 4 genes,
XX EBL-rel (AA083526), E3la (AA083527), EBL-e2 (AA083528) and Proj3
XX (AA083529), encode the proteins given in AAR70233-36, respectively. The
XX binding domains of such proteins can be expressed e.g. in E. coli,
XX yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
XX cells, and provide protection against P. falciparum.
XX
XX Sequence 700 AA;
XX
XX Query Match 5.3%; Score 1032.5; DB 16; Length 700;
XX Best Local Similarity 34.8%; Pred. No. 2e-53;
XX Matches 247; Conservative 110; Mismatches 250; Indels 103; Gaps 22;
OY 136 NDALCAPRRRHODKLEALINDINQINDILGNVLYTKYGESESVNHH--PHGCT- 192
DB 5 NKVGAACAYRRLHLCIDVLESID--TITTKLLEVCMAAKYBNSI-NHYHQHQTN 61
OY 193 ----SDACTALARSFADIGDIVRGIDMF-----KPNVHDKVETGLREVFRIHGMDE 242
DB 62 BDSASQCLCTVLARSFADIGDIVRGIDVLYGYDNKEBQKRLKQKIDIFKIH---KDV 118
OY 243 VKND-----VMPDG-SGVYKTLREAMNVMNKNKWEALITCJASKSGTFMOSNTPLFS 296
DB 119 MKTNGAQRERTIDDAKGDFQLREDQWTSNREYWKALICHAPPEANFTITACNVGKT 178
OY 297 NPKGSHKQKVPNTLDYVPOYLRFWDEWEGFCKRRNKLTKKVDSCNDKERLYCSHG 356
DB 179 NGQCHICIGDVPYTFDYVPOYLRFWEBAEDFCRKKKLENLQCKRDYQNLKCSGNG 238
OY 357 HDCTTTKKGILHLDNKCTDCSTKCKVEYVWLGNOQEAFFKQREYKEI----- 407
DB 239 YDCRTKYTKKGLIVGEHCTNCVWCRMYETWIDNOKREFLKQKRYETELISGGGSKSP 298
OY 408 -----QSLSDNKNFVNINSEYKQFYEKAKETQYATANDFFLNLNDEGTYCKG--- 456
DB 299 KRTKRAARSSSSSDN-----GYESKFEYKLLKEVGYODVDKFLILNKEGICQKOPQ 350
OY 457 -GLPGEKIDFTFTNSADKGIYRSEYCOVCPGCVKCGDGIKYTRKSDNDRVRNNEYKP 515
DB 351 VGNKEKADNVDFTNKRYK-FTSRTEICEPCWCLEGGKGPWKYKSGDKTGSATKTYDP 409
OY 516 PWGVKPTNITVLYSGNREGDITQKLENFCSNSTNYKDKKNQKWEYCYKD-----ENINRC 570
DB 410 K--NITDIPVLYPDKSOONILKRYKNCCEGAP-GGQIWKQCYCYDEHRPSSKNNNC 465
OY 571 KLENTLEINDNPNKIIISFHNFEELWYTYLLBDTIKMDKLTCTCINNNTT---HCIDECN 626
DB 466 VEGTWKDFTOGKQYKSYNVFMDVHDMKLDHSEVEMTELSKCIINNNTNGTCRNKKK 525
OY 627 RNCLCFDRWYKQKEEWNISIKKLTFRKKNITQOASYYSININLFEYGFYKVDKLDKDAKW 686
DB 526 TDCGCFQKWEYKQOEWMAIKDHGKQTDLYQO-----KGLIYSPYGVLDIV---LKG 576

```

QY 687 KEMENIKRKNENLNNDYLE-----NAIELLDHLKETATICKDNN 732
 DB 577 GNLQNIKDVHGDYDDIKIKIKLLDEEDAVAVYLGKDNNTIDIKLQHEKQOAEQCKOK- 635
 OY 733 TNEACE-----TSHNATTNPKVPRGTOPTKNIKEIAOYFRSAAYEE 775
 DB 636 -QEECEKKAQOESRGRSAETREDERTQ---QPADSAGEVEEEDDDDDYDE 681

RESULT 15
 AAM22481

ID AAM22481 standard; protein; 700 AA.

XX AC AAM22481;

DT 07-OCT-1997 (first entry)

XX DE Plasmodium ebl-2.

DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
 Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 Plasmodium.

XX OS Plasmodium falciparum.

PN W09640766-A2.

PD 19-DEC-1996.

XX PF 07-JUN-1996; 96MO-US09508.

XX PR 07-JUN-1995; 950S-0487826.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PI Chittis C, Miller LH, Peterson DS, Sim KL, Su X;

XX PI Wellens TE;

XX DR WPI: 1997-052231/05.

XX DR N-PADB: AAT72896.

XX PT New malaria vaccines - contains cysteine-rich DBL family protein
 binding domains homologous domains of the Duffy and sialic acid
 binding proteins

XX PS Disclosure: Page 46-48; 96pp; English.

XX This sequence represents ebl-2 of Plasmodium. Ebl-2 belongs to
 the Duffy binding like (DBL) family of genes which have homology to the
 Duffy antigen binding protein (DABP) and sialic acid binding protein
 (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 var family of genes modulate cytoadherence and antigenic variation of
 Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 protein (DABP) are soluble proteins that appear in the culture
 supernatant after infected erythrocytes release merozoites. DABP and
 SABP mediate the binding of merozoites and schizonts to the erythrocyte
 surface. These proteins are necessary for erythrocyte invasion by the
 parasite. This sequence can be used in the compositions of the
 invention. The compositions are for the treatment and prevention of
 malaria, and comprise either a nucleotide sequence or encoded polypeptide
 of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
 family of genes having homology with conserved regions of DABP and SABP.
 The compositions are used for the treatment and prevention of malaria.
 They are also used in the preparation of vaccines for inducing a
 protective immune response in a mammal to Plasmodium merozoites
 (especially Plasmodium falciparum or Plasmodium vivax).

XX SQ Sequence 700 AA;

Query Match 5.3%; Score 1032.5; DB 18; Length 700;
 Best Local Similarity 34.8%; Pred. No. 2e-53;
 Matches 247; Conservative 110; Mismatches 250; Indels 103; Gaps 22;

QY 136 NDATACAPRRRHKCDKNLEALNDINTQNIHDLGNLVYAKYEGESIYNNH--PHKGT- 192
 DB 5 NKVACAPRYRLHLCIDYNLSEID--TSTTHKILLLEVCMAKYAGNSI-NHYHQHRTN 61
 QY 193 ----SDACTALARFPAIDGIVRGIDMF-----KPNVHDKVEFGLDEYFKKHIDGMEDE 242
 DB 62 EDSASQLCTVIALRSEFAIDGIVRGIDLYLGYNKEKBORRKLQKLDIFKRIH---KDV 118
 QY 243 VKND-----YNPDG-SGNYYKLRAMNNVNRNRYEAITCDASYSQYFMOSESTPLFS 296
 DB 119 MKTNGAERIYIDDAKGDGDFQLRREDMWT SNREYWKALICHAPKEANYFTKACNVGKT 178
 QY 297 NPKCGHKGAVPTNLDYVPOYLRFWDEMGEEFCRRRIKIKRVKDSGRNDRERLYCSHG 356
 DB 179 NGQCHICIGDVPPTFYFDYVPOYLRFWEEMAEDEFCRKKKKLENILOKQCRDYEQNLCSGNG 238
 QY 357 HDCTTWMKKGIHLMDKCKDCSTKCYFEFVWLGNOQEAFAKKQEKEXEKEI----- 407
 DB 239 YDCTTLYKRGKLVIGHCSTGYSWCRMYETWIDNOKREPLKORRKYETEISGGSGSKSP 298
 QY 408 -----OSYLSNDNRFVNNINSEYKOFYEKLETOYATNDPFLNLENGKYCKG--- 456
 DB 299 KRTKRAARSSSSSDN-----GYESKFYKIKLEYGYQDVDFLKIINKEGICQKOPQ 350
 QY 457 -GLPEKDIPTFTNSADKGIYRSEYQVCPDCGVKDGKIKYTHKSDNDRERVNNEDYKP 515
 DB 351 VGNEKADNVDTNNKRYK-FESRTEICEPCPWCGLKRGKGPWKYKGDTCGSAKTKYDP 409
 QY 516 PMGVKPNINIVLYSGNQGDITOKLEFNCNSTYKXKNKNNKWCYKXD-----ENINRC 570
 DB 410 K---NITDIPLVYDKSQONILKRYKNFCEKAGP-GGGQIKKWQCYDDEHRPSSKNNNC 465
 QY 571 KLEQTEINNDNPKIISFHNFEFLWYTLRDTIKMNDKLTCTCINNTT-----HCIDECN 626
 DB 466 VEGTWDFKFTQCKQVYKSYNFFMDVYHDLDSVEMTELKSCJNNNTNGTCRNNNCK 525
 QY 627 RNCICFDRAWYKQKEENNSIKLFTKRNKNIQOOSYYSNINLFBGYFPKVMKDLKDEAKW 686
 DB 526 TDCGCFQKWEKQOEWMAIKDHFQKOTDIVQO-----KGLIVSPYGVLDLV-----LKG 576
 QY 687 KEMENIKRKNENLNNDYLE-----NAIELLDHLKETATICKDNN 732
 DB 577 GNLQNIKDVHGDYDDIKIKIKLLDEEDAVAVYLGKDNNTIDIKLQHEKQOAEQCKOK- 635
 QY 733 TNEACE-----TSHNATTNPKVPRGTOPTKNIKEIAOYFRSAAYEE 775
 DB 636 -QEECEKKAQOESRGRSAETREDERTQ---QPADSAGEVEEEDDDDDYDE 681

Search completed: April 28, 2003, 10:29:57
 Job time : 224.055 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 132.729 Seconds

(without alignments)
5498.552 Million cell updates/sec

Title: US-10-087-013-2

Perfect score: 19407
Sequence: 1 MGESCKYFIKMGNAASSLE.....IDLINDSLVNLILLYMKY 3542

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

1 number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19407	100.0	3542	5	Q9U5M2 plasmodium
2	12954.5	66.8	3287	5	Q8T326 plasmodium
3	7367	38.0	1615	5	Q8T325 plasmodium
4	4515.5	23.3	2706	5	Q15870 plasmodium
5	4254	21.9	3006	5	Q26032 plasmodium
6	4010	20.7	3026	5	Q26030 plasmodium
7	3489	18.0	3078	5	Q26031 plasmodium
8	3328.5	17.2	2527	5	Q95W83 plasmodium
9	3130	16.1	2924	5	Q25733 plasmodium
10	2932	15.1	1685	5	Q9U4A2 plasmodium
11	2869	14.8	2658	5	Q8T360 plasmodium
12	2817	14.5	510	5	Q8T6T0 plasmodium
13	2714.5	14.0	2664	5	Q26033 plasmodium
14	2686.5	13.8	2647	5	P90580 plasmodium
15	2677	13.8	2212	5	Q94657 plasmodium
16	2556.5	13.2	494	5	Q8T6K7 plasmodium

17	2480.5	12.8	492	5	Q8T6K9 plasmodium
18	2449.5	12.6	494	5	Q8T6K0 plasmodium
19	2441.5	12.6	465	5	Q8T6K4 plasmodium
20	2436	12.6	2710	5	Q9XZB8 plasmodium
21	2413.5	12.4	494	5	Q8T6K5 plasmodium
22	2412.5	12.4	492	5	Q8T6K6 plasmodium
23	2403.5	12.0	496	5	Q8T6K8 plasmodium
24	2335.5	12.0	461	5	Q8T6K3 plasmodium
25	2303	11.9	2135	5	Q61077 plasmodium
26	2258.5	11.6	460	5	Q8T6K1 plasmodium
27	2056.5	10.6	1327	5	Q9NFB4 plasmodium
28	2055.5	10.6	427	5	Q8T6K2 plasmodium
29	1878.5	9.7	2169	5	Q97312 plasmodium
30	1750	9.0	2163	5	Q9NFB6 plasmodium
31	1749	9.0	2197	5	Q96296 plasmodium
32	1743	9.0	2182	5	Q26034 plasmodium
33	1716.5	8.8	2209	5	Q97324 plasmodium
34	1697.5	8.7	2227	5	Q9U0G5 plasmodium
35	1689.5	8.7	2228	5	Q60991 plasmodium
36	1611	8.3	2209	5	Q9U0G6 plasmodium
37	1574	8.1	1711	5	Q96108 plasmodium
38	1523.5	7.9	2042	5	Q25766 plasmodium
39	1249	6.4	1729	5	Q25734 plasmodium
40	924	4.8	173	5	Q96735 plasmodium
41	811.5	4.2	431	5	Q96294 plasmodium
42	749.5	3.9	2647	5	Q9U4X0 plasmodium
43	660.5	3.4	438	5	Q9Y1N7 plasmodium
44	651	3.4	118	5	Q96450 plasmodium
45	646.5	3.3	455	5	Q9Y1N6 plasmodium

ALIGNMENTS

RESULT 1

Q9U5M2 PRELIMINARY; PRT; 3542 AA.
AC 09U5M2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2002 (TREMBLrel. 20, Last annotation update)
DE PCR3 CSA ligand (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCR3;
RX MEDLINE=2006305; PubMed=10535993;
RA Buffet P., Gamain B., Scheldig C., Baruch B., Oishi S., Fujii N.,
Fusai T., Paray D., Miller L.H., Gysin J., Scherf A.;
RT "Plasmodium falciparum domain mediating adhesion to chondroitin
sulfate A: A receptor for human placental infection.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:12743-12748(1999).
DR EMBL; AJ133811; CAB59840.1; -
DR InterPro; IPR001219; Neurotoxin.
DR Pfam; PF03011; PREMP. 1.
DR PRINTS; PR00284; TOXIN.
FT NON_TER 3542
SQ SEQUENCE 3542 AA; 413089 MW; 970D85EB8BDA2BC2 CRC64;

Query Match 100.0%; Score 19407; DB 5; Length 3542;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3542; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGESCKYFIKMGNAASSLEGDASKSPIIKESHKSARNVLEARYAKNIRHPSYAKAEHVDL 60
DB 1 MGESCKYFIKMGNAASSLEGDASKSPIIKESHKSARNVLEARYAKNIRHPSYAKAEHVDL 60
QY 61 KGDITAEFRGGRSPYVKNHNYTPPCULDKRHNLAQYDVNLNHPCHGRQGNFDD 120
|||||

Db 61 KGLTLAEFRGSGSTPVNKNHNYYPYPCNLDHKEHTNLRYDYNLRHPRGREGONRDE 120
Qy 121 ESECGKINRYKRNDAICAPRRRHMCDDKULEALNDINTONIDILGNVYTAKEYS 180
Db 121 ESECGKINRYKRNDAICAPRRRHMCDDKULEALNDINTONIDILGNVYTAKEYS 180
Qy 181 ESTVNNHPRHGTSADCTALARSFADIGDLYRGIDMFKPNVNDKVEYGLREVEFKIHGME 240
Db 181 ESTVNNHPRHGTSADCTALARSFADIGDLYRGIDMFKPNVNDKVEYGLREVEFKIHGME 240
Qy 241 DEKNDYNDPGSGNYYKLRAMNVRNRYWEAITCDASYKSYEMQSSNPTLPFSNPKC 300
Db 241 DEKNDYNDPGSGNYYKLRAMNVRNRYWEAITCDASYKSYEMQSSNPTLPFSNPKC 300
Qy 301 GHRQGVPTNLDYVQYLRWFEDMEGECRKRNIKLKYKDCSRNKEKELYCSHNGHDT 360
Db 301 GHRQGVPTNLDYVQYLRWFEDMEGECRKRNIKLKYKDCSRNKEKELYCSHNGHDT 360
Qy 361 TTIWKKGIILHDKKCDSCSTKCVFEVWLGNOEAKKOKEREKEIOYLSNDNKFVNN 420
Db 361 TTIWKKGIILHDKKCDSCSTKCVFEVWLGNOEAKKOKEREKEIOYLSNDNKFVNN 420
Qy 421 INSEYKQFYEKLKETQYATNDTFLNLNEKRYCKGLGEGEDITFTNSADOKGIFYSSE 480
Db 421 INSEYKQFYEKLKETQYATNDTFLNLNEKRYCKGLGEGEDITFTNSADOKGIFYSSE 480
Qy 481 YQOVPDGGVCKDGYKYTHKSDNDEBRVANNEDYKPPWGVKPTNTIVLYSGNEQDITOKL 540
Db 481 YQOVPDGGVCKDGYKYTHKSDNDEBRVANNEDYKPPWGVKPTNTIVLYSGNEQDITOKL 540
Qy 541 ENFCSSSTNYKDNOKNOKWECYKEDENINCKLEONTLEINNDPKIISPHNPELWVYLL 600
Db 541 ENFCSSSTNYKDNOKNOKWECYKEDENINCKLEONTLEINNDPKIISPHNPELWVYLL 600
Qy 601 RBTIMWNLKKTICINNTTTHCIDECONRCLCPDRWVYKOKEEBWSIKKLFTRKKNIOOSY 660
Db 601 RBTIMWNLKKTICINNTTTHCIDECONRCLCPDRWVYKOKEEBWSIKKLFTRKKNIOOSY 660
Qy 661 YSNINLEGEYEFKYMDLKDDEAKWKELEMEIKRKKNEFSNLENRNYLENAIEELLDH 720
Db 661 YSNINLEGEYEFKYMDLKDDEAKWKELEMEIKRKKNEFSNLENRNYLENAIEELLDH 720
Qy 721 LKETATICKDNNTNACETSHNATTPNCVPRGTOPTKNTKEIAQYFKRSAYEARNRG 780
Db 721 LKETATICKDNNTNACETSHNATTPNCVPRGTOPTKNTKEIAQYFKRSAYEARNRG 780
Qy 781 LHKLAGRAHEGITYRGRGRKDKNLCRIYKHSNRNLGFSNGCQDGKGTGDIQTRFVY 840
Db 781 LHKLAGRAHEGITYRGRGRKDKNLCRIYKHSNRNLGFSNGCQDGKGTGDIQTRFVY 840
Qy 841 GTEWEVDEHMRKDHEDYMPRRRHICTSNLEHLOTDDHPLNINYDIDLNNSEFLDYL 900
Db 841 GTEWEVDEHMRKDHEDYMPRRRHICTSNLEHLOTDDHPLNINYDIDLNNSEFLDYL 900
Qy 901 LSAAYEANKIIRMYKKNKLGPREVTDPKHQTTCIRAIRYSFADIGDIIRGRDLWRNG 960
Db 901 LSAAYEANKIIRMYKKNKLGPREVTDPKHQTTCIRAIRYSFADIGDIIRGRDLWRNG 960
Qy 961 DMVLAQHLLETFVGNHKSJLKGKNDKYNDAPRYLKRNNWEANAKWEMAKCDIKY 1020
Db 961 DMVLAQHLLETFVGNHKSJLKGKNDKYNDAPRYLKRNNWEANAKWEMAKCDIKY 1020
Qy 1021 LKRSQHSQSTQSGYSDHTPLDDYIPQRLRMTEWAEWYCYVQKKEYDKLEKCECK 1080
Db 1021 LKRSQHSQSTQSGYSDHTPLDDYIPQRLRMTEWAEWYCYVQKKEYDKLEKCECK 1080
Qy 1081 DKDNGOCTSESGTCKCEACNEVNDIIGLKEQMNISDKYKELHBOQMSVSNST 1140
Db 1081 DKDNGOCTSESGTCKCEACNEVNDIIGLKEQMNISDKYKELHBOQMSVSNST 1140
Qy 1141 EASSTAKNHIDRVIEFLSELGYOONGSKSNKSGTSDSAVIGTNTYENYNGALHTGTGNF 1200
Db 1141 EASSTAKNHIDRVIEFLSELGYOONGSKSNKSGTSDSAVIGTNTYENYNGALHTGTGNF 1200

Qy 1201 DDCQSONEFCDEKSDGKDNKEKAFPRDKPODDHAGACGCKSGSKPTRYOIKTKKAAEKDTE 1260
Db 1201 DDCQSONEFCDEKSDGKDNKEKAFPRDKPODDHAGACGCKSGSKPTRYOIKTKKAAEKDTE 1260
Qy 1261 CXTVNDIILEKNGKQVDECHKKNSNGYPRMOGGINLVEDPRVCMPPRQKLCVHFLA 1320
Db 1261 CXTVNDIILEKNGKQVDECHKKNSNGYPRMOGGINLVEDPRVCMPPRQKLCVHFLA 1320
Qy 1321 NDNEIKKLOSONYKLEAFIKSAAETFFSWTYYSKQEGEHELKELKEGTPAPFLRSM 1380
Db 1321 NDNEIKKLOSONYKLEAFIKSAAETFFSWTYYSKQEGEHELKELKEGTPAPFLRSM 1380
Qy 1381 FYTBEDYDFLPGTIDISGHEGSKLEQIDSLFRKNGDOKSPNGKTRQEWMTESHEIWE 1440
Db 1381 FYTBEDYDFLPGTIDISGHEGSKLEQIDSLFRKNGDOKSPNGKTRQEWMTESHEIWE 1440
Qy 1441 AMLCALVIGAKKDPFENYVNNVKSFSKSTLEEFRAKRPQFLRWLTWYDYCYTRQK 1500
Db 1441 AMLCALVIGAKKDPFENYVNNVKSFSKSTLEEFRAKRPQFLRWLTWYDYCYTRQK 1500
Qy 1501 YLKDVOEKCSNDOLKCDTECNKCEDEYVYKMKKKEMIFPODKYKPERDKRFRDQHG 1560
Db 1501 YLKDVOEKCSNDOLKCDTECNKCEDEYVYKMKKKEMIFPODKYKPERDKRFRDQHG 1560
Qy 1561 VMATDYGTNATDYLNKRFTASGDRKPGSASVYQVRIQLLEKQAYYDADKHCCCTYFEN 1620
Db 1561 VMATDYGTNATDYLNKRFTASGDRKPGSASVYQVRIQLLEKQAYYDADKHCCCTYFEN 1620
Qy 1621 DDKYTNISSKDKCKGLVKEANTGAIKQONKGPNNYNNLKELTBDYVLPSPRLICHAID 1680
Db 1621 DDKYTNISSKDKCKGLVKEANTGAIKQONKGPNNYNNLKELTBDYVLPSPRLICHAID 1680
Qy 1681 GNTYDPEVOKENGIRKRLMVAATTEGYNLCQYKREKKEKKTITSADHKYSYEVPCSAM 1740
Db 1681 GNTYDPEVOKENGIRKRLMVAATTEGYNLCQYKREKKEKKTITSADHKYSYEVPCSAM 1740
Qy 1741 KYSEFYDLRDIILIGIDNLEDEKQTEENLKKIFPNKNGTSVGKSGSDSTGPGSTARFEFN 1800
Db 1741 KYSEFYDLRDIILIGIDNLEDEKQTEENLKKIFPNKNGTSVGKSGSDSTGPGSTARFEFN 1800
Qy 1801 ENKECVNANMTCYKKGRODNGSNGSARSDEDLKKGCSVPSDDYPMGNRDEGTAYOFL 1860
Db 1801 ENKECVNANMTCYKKGRODNGSNGSARSDEDLKKGCSVPSDDYPMGNRDEGTAYOFL 1860
Qy 1861 RMEFNGEDFCRKEKELEKLVAGANDYTCGDNEDKRRKCTACTQYKFFISEMRPOYER 1920
Db 1861 RMEFNGEDFCRKEKELEKLVAGANDYTCGDNEDKRRKCTACTQYKFFISEMRPOYER 1920
Qy 1921 QIKKYEENKDKIYSEHPVAKDAEDAREYLDKOLKICEKNSGDCYKCKMDYSTQRLTDG 1980
Db 1921 QIKKYEENKDKIYSEHPVAKDAEDAREYLDKOLKICEKNSGDCYKCKMDYSTQRLTDG 1980
Qy 1981 NSQNMPSASIDDEPKVEGKCNCOVPRGPRVRETPSPRVSLSKATASKEAKTAPPTK 2040
Db 1981 NSQNMPSASIDDEPKVEGKCNCOVPRGPRVRETPSPRVSLSKATASKEAKTAPPTK 2040
Qy 2041 QPKYVENLTTEBRAQTRPRRAQOTRKTSTATTETESDVGTVKAIILSNKPSRSGTSC 2100
Db 2041 QPKYVENLTTEBRAQTRPRRAQOTRKTSTATTETESDVGTVKAIILSNKPSRSGTSC 2100
Qy 2101 NPRTYGOYKWCICYGKSENGNIGCMPPRRKKLCINNIQYLYNTEENKRDNDIKEAFIK 2160
Db 2101 NPRTYGOYKWCICYGKSENGNIGCMPPRRKKLCINNIQYLYNTEENKRDNDIKEAFIK 2160
Qy 2161 CAEIFOFLMILKYIENPAEENELONGTIPDEFRKIMYYTGVDYKDMFRGDISNDKKIT 2220
Db 2161 CAEIFOFLMILKYIENPAEENELONGTIPDEFRKIMYYTGVDYKDMFRGDISNDKKIT 2220
Qy 2221 TYTNSVTTIILNENKKNODKDEELARKIFWEKKNKFTMEGNIIGLYLHLDENEKEKIR 2280
Db 2221 TYTNSVTTIILNENKKNODKDEELARKIFWEKKNKFTMEGNIIGLYLHLDENEKEKIR 2280

QY 2281 DNYQYDMTKLPSPLEEFYKRRPOFLMFTENAEFCNKRREOLKLEAGCKEYECGNSD 2240
 DB 2281 DNYQYDMTKLPSPLEEFYKRRPOFLMFTENAEFCNKRREOLKLEAGCKEYECGNSD 2240
 QY 2341 GTOCEACACVYQNFITKMKTEYERQOREKFKDKGKKYKDYPTSERDIEKATCAHEYL 2400
 DB 2341 GTOCEACACVYQNFITKMKTEYERQOREKFKDKGKKYKDYPTSERDIEKATCAHEYL 2400
 QY 2401 NMLKLELCSNKCSCQMKPSSOLPKTTQOSQSDANDMPESLDLYPEEFNCKCEPESLKR 2460
 DB 2401 NMLKLELCSNKCSCQMKPSSOLPKTTQOSQSDANDMPESLDLYPEEFNCKCEPESLKR 2460
 QY 2461 GSWIHFKKTIETERTIPANCYKCAAYILSKAENMDITLKEKFIPISTYKESKNSWTNN 2520
 DB 2461 GSWIHFKKTIETERTIPANCYKCAAYILSKAENMDITLKEKFIPISTYKESKNSWTNN 2520
 QY 2521 NPODPPKPYAPDKYIGRRNPPCENRENRKVDYEMKCYNKSFKYOKRKRVCPPRRHMC 2580
 DB 2521 NPODPPKPYAPDKYIGRRNPPCENRENRKVDYEMKCYNKSFKYOKRKRVCPPRRHMC 2580
 QY 2581 LRLNDEIKTERLKSNDVYLKMWRTARNEGIDILKFNPSNGCAMPICDTMKYSFADLG 2640
 DB 2581 LRLNDEIKTERLKSNDVYLKMWRTARNEGIDILKFNPSNGCAMPICDTMKYSFADLG 2640
 QY 2641 DIVRGTDMLRIGGYLPVEIKLYVEYIYGKRNKNGKRNKYNDVQTERSAMDANRKO 2700
 DB 2641 DIVRGTDMLRIGGYLPVEIKLYVEYIYGKRNKNGKRNKYNDVQTERSAMDANRKO 2700
 QY 2701 IMKAMTCAPEDAKLFRRKGMOGEFRTLLIODCGHKDPPVDYIIPORRMWTESEXY 2760
 DB 2701 IMKAMTCAPEDAKLFRRKGMOGEFRTLLIODCGHKDPPVDYIIPORRMWTESEXY 2760
 QY 2761 CRALMELEKFKKSCDCHTSRCKNDYDENKCEQCKTRQOEKKNVYLKWSLFDIOSNK 2820
 DB 2761 CRALMELEKFKKSCDCHTSRCKNDYDENKCEQCKTRQOEKKNVYLKWSLFDIOSNK 2820
 QY 2821 YKELYQPIYTKISTYDHOYQYQAKLTKRSECVSEFSFSEYTHETSKCLNFKENNDGSS 2880
 DB 2821 YKELYQPIYTKISTYDHOYQYQAKLTKRSECVSEFSFSEYTHETSKCLNFKENNDGSS 2880
 QY 2881 NRTYAFEEPTPKSYKACSTLPSKNPLDNCPLDOKKDGCKELQOTPFCSKNYDNNLND 2940
 DB 2881 NRTYAFEEPTPKSYKACSTLPSKNPLDNCPLDOKKDGCKELQOTPFCSKNYDNNLND 2940
 QY 2941 WNAVYLVNSSDOKGVLIPRRRHLCRPTTAYNRRGDEILKRLTSLAFSOGOLLG 3000
 DB 2941 WNAVYLVNSSDOKGVLIPRRRHLCRPTTAYNRRGDEILKRLTSLAFSOGOLLG 3000
 QY 3001 KYKSEELCFEAMKYADYSDILKCTDMMDISLSEKIKKIFETSNEATEENRKTWENNRR 3060
 DB 3001 KYKSEELCFEAMKYADYSDILKCTDMMDISLSEKIKKIFETSNEATEENRKTWENNRR 3060
 QY 3061 ROJWHAHLGKYATATSKVTLDEGMCOLPDEETNOFLRMLIEMAKACRKKKRVDSILKT 3120
 DB 3061 ROJWHAHLGKYATATSKVTLDEGMCOLPDEETNOFLRMLIEMAKACRKKKRVDSILKT 3120
 QY 3121 KCPRSNEDNEASELLROPGCONDIRKYISLNLINKNTMENLILIKYKOLKDOSSGNIDNK 3180
 DB 3121 KCPRSNEDNEASELLROPGCONDIRKYISLNLINKNTMENLILIKYKOLKDOSSGNIDNK 3180
 QY 3181 PSEENQSYTKSKDSCALELNDINEIVGTKNENNEFEVYKALYPCLYEVEDETHKN 3240
 DB 3181 PSEENQSYTKSKDSCALELNDINEIVGTKNENNEFEVYKALYPCLYEVEDETHKN 3240
 QY 3241 HVLGDNITKEEOTVRKALYFTPHVDSFYQAPLFTSHRYAQYDPKNDILKSSISVYIYS 3300
 DB 3241 HVLGDNITKEEOTVRKALYFTPHVDSFYQAPLFTSHRYAQYDPKNDILKSSISVYIYS 3300
 QY 3301 ALGLIALHFKKKKSSVDLRLTILNIPQGEYGMPTLESKNRYIPYSGPYKGGATYIYMEG 3360
 DB 3301 ALGLIALHFKKKKSSVDLRLTILNIPQGEYGMPTLESKNRYIPYSGPYKGGATYIYMEG 3360
 QY 3361 DTSGDEDKYMWDLSSDITSSSEYEELDINDIYVPGSPKYTLIEVYLEPSKRDIIPSD 3420

DB 3361 DTSGDEDKYMWDLSSDITSSSEYEELDINDIYVPGSPKYTLIEVYLEPSKRDIIPSD 3420
 QY 3421 TPSENDPRTNRFFIDDEMNELKHPVSOYLPENTEPNNNYKCAADIPMTEPTTLSDNPEEK 3480
 DB 3421 TPSENDPRTNRFFIDDEMNELKHPVSOYLPENTEPNNNYKCAADIPMTEPTTLSDNPEEK 3480
 QY 3481 PFITSIHNDLYTGKEISYNINNSTNTNNDIPMNRNDSYRGIDLINDSLVNLILYMM 3540
 DB 3481 PFITSIHNDLYTGKEISYNINNSTNTNNDIPMNRNDSYRGIDLINDSLVNLILYMM 3540
 QY 3541 KY 3542
 DB 3541 KY 3542
 RESULT 2
 ID 08T326 PRELIMINARY; PRT; 3287 AA.
 AC 08T326;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
 DE PFEMP1 (Fragment).
 GN TW180VAR2.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TW180;
 RX MEDLINE=21927235; PubMed=11930336;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT Identification of a conserved Plasmodium falciparum var gene
 RT implicated in malaria in pregnancy.";
 RL J. Infect. Dis. 185:1207-1211(2002).
 DR EMBL; AJ20411; CAD20867.1; -;
 FT NON_TER 3287 3287
 SQ SEQUENCE 3287 AA; 383550 MW; 58F8E866FC244536 CRC64;
 Query Match 66.8%; Score 12954.5; DB 5; Length 3287;
 Best Local Similarity 72.7%; Pred. No. 0;
 Matches 2440; Conservative 267; Mismatches 524; Indels 125; Gaps 43;
 QY 12 MGAASLEBDASPIIKESHKSARNVLEKRNIRPSYAKEHVDSLKGDTLKAFFRG 71
 DB 1 MGAASLEBDASPIIKESHKSARNVLEKRNIRPSYAKEHVDSLKGDTLKAFFRG 60
 QY 72 GPSTPVNKHNYYPYPCNLDHKEHTNLRYDVLNRHPCHGREGNREDESEECGNKIRN 131
 DB 61 GPSTPVNKHNYYPYPCNLDHKEHTNLRYDVLNRHPCHGREGNREDESEECGNKIRN 120
 QY 132 YKRNDAIACAPRRRRMCKNLLEALNDITQNTIHDGLGVLYTAKYEGSIYNNHHRK 191
 DB 121 YKRNDAIACAPRRRRMCKNLLEALNYINTQNIHDGLGVLYTAKYEGSIYNNHHRK 180
 QY 192 TSDACTALASFPDIDGIVGIDMFKNPVHDKYETGLREYFKKIHDMCEBEVKNYVDPG 251
 DB 181 TSDVCTALASFPDIDGIVGIDMFKNPVHDKYETGLREYFKKIHDMCEBEVKNYVDPG 240
 QY 252 SGNYTKILREAMVNNRNKRVWEATITCDASYKSGYFMOSSESPTLPSNKGCH-KGGVPTN 310
 DB 241 SGNYTKILREAMVNNRNKRVWEATITCGALPKSAVFLQSEDNKOFLVLPKCGHNNKNDLPTN 300
 QY 311 LDVVPOLRFEDMGSEFCRKNITKLKVVDSORNKERYLCSHGNGDCTTLMKGLILH 370
 DB 301 LDVVPOLRFEDMGSEFCRKNITKLKVVDSORNKERYLCSHGNGDCTTLMKGLILH 360
 QY 371 LDNRCTDCSTYKVFYEWLGNQOEAFKKQKEKEKEIOSYLSNDNKNVNNINSEYKQFY 430
 DB 361 LDNRCTDCSTYKVFYEWLGNQOEAFKKQKEKEKEIOSYLSNDNKNVNNINSEYKQFY 420
 QY 431 EKLKETOYATNTFTLNLNLENGKYCKGGLPGEKDITFTNSADKGIIFYRSBYCOVCPDCGV 490

```

Db 421 DDLDDKNNHNDLDTPLNLINLEKGYCKEKIKGESDITFTNSSDDKGIFFRSQYCYQCPDGGV 480
Qy 491 KCDIKYTHKSDNREHREVNEDYKPPMGVAPPTNTTVYLSGNEGDLTOXKLENCNSSTN 530
Db 481 KCDGTATYTHSDNRECVNEDYKPPMGVAPPTNTTVYLSGNEGDLTOXKLENCNSSTN 540
Qy 551 KDKNNQWECYKQDENINRCKLEONTEINNDNPKIISFHFNFELWYLYLDRTIKWMDL 610
Db 541 KDKNNQWECYKQDENINRCKLEONTEINNDNPKIISFHFNFELWYLYLDRTIKWMDL 600
Qy 611 KTCINNTTICIDECNNGCICPDWYVYKQKEEENSISIKLETKKNNQOOSTYSININLEFG 670
Db 601 KTCINNTTICIDECNNGCICPDWYVYKQKEEENSISIKLETKKNNQOOSTYSININLEFG 660
Qy 671 YFFKVMKDLDKBAKWELEMINIKRKKNFESNLENNRDYLENAIELLDLTKETATCKD 730
Db 661 YFFKVMKDLDKBAKWELEMINIKRKKNFESNLENNRDYLENAIELLDLTKETATCKD 720
Qy 731 NNTNNECETSHNATITPVCVAPRGCTOPTKNIKIAOYFKRSAYEARNRGLHKLKRAHE 790
Db 721 NNTNNECETSHNATITPVCVAPRGCTOPTKNIKIAOYFKRSAYEARNRGLHKLKRAHE 780
Qy 791 GIYKRGRRKDFKDNLCRIKIKHSNNRNLGFSNGCDGKGIOIRFYVGTWEDPEH 850
Db 781 GIYKRGRRKDFKDNLCRIKIKHSNNRNLGFSNGCDGKGIOIRFYVGTWEDPEH 840
Qy 851 MKRQHDVYIMPPRRRHIGTSNLEHLOTDDHPLNGNIVDVLVNSFLGVDLLSAKYANKI 910
Db 841 MKRQHDVYIMPPRRRHIGTSNLEHLOTDDHPLNGNIVDVLVNSFLGVDLLSAKYANKI 900
Qy 911 IRMYKKNLKGPKKEDYDPRKHOTTCRATRSFADIGDIIRGRDLMERNGDWYKLGHE 970
Db 901 IRMYKKNLKGPKKEDYDPRKHOTTCRATRSFADIGDIIRGRDLMERNGDWYKLGHE 960
Qy 971 YVFGNTHSLKAGNDKYNDAPKYLKLENNWEANRAKWEAMKCDIYLLKDKSGHST 1030
Db 961 YVFGNTHSLKAGNDKYNDAPKYLKLENNWEANRAKWEAMKCDIYLLKDKSGHST 1020
Qy 1031 QSSYCSYSHTEPLDYIIPQKLMWTEWAMWYCKVOKEYDKLEKCKECKDKONGQCTK 1090
Db 1021 QSSYCSYSHTEPLDYIIPQKLMWTEWAMWYCKVOKEYDKLEKCKECKDKONGQCTK 1080
Qy 1091 ESGTCTCTEACNEYNNDIIGLMEOMNIIISDYKELHEOAOMSVNSGIESSTAKNHI 1150
Db 1081 ESGTCTCTEACNEYNNDIIGLMEOMNIIISDYKELHEOAOMSVNSGIESSTAKNHI 1140
Qy 1151 DRYVIEFLSELYOONGSKNSKSGTSDGAVIGTNTTYENVGAYLHDGTFDQOSONEFC 1210
Db 1141 DRYVIEFLSELYOONGSKNSKSGTSDGAVIGTNTTYENVGAYLHDGTFDQOSONEFC 1200
Qy 1211 DEKSDGKNEKAFPRDKODHDGAGCGSGSPRTVOIKTKKAKBEKDECKEYVNDILKE 1270
Db 1201 DEKSDGKNEKAFPRDKODHDGAGCGSGSPRTVOIKTKKAKBEKDECKEYVNDILKE 1260
Qy 1271 NDGKQVEDCHPRKNSNGYPRMOC---GNINLVEDPRVAPPRRORCKVAFIANDNEIK 1326
Db 1261 NDGKQVEDCHPRKNSNGYPRMOC---GNINLVEDPRVAPPRRORCKVAFIANDNEIK 1316
Qy 1326 NGGTJKVGEVNCVKTGE-YPEWCECKKQONH---KACMPPRPKLCVAFIHE---L 1394
Db 1316 NGGTJKVGEVNCVKTGE-YPEWCECKKQONH---KACMPPRPKLCVAFIHE---L 1304
Qy 1394 KLOSGVNLKEAFISAAETFFSWYYSKDGEGNELDKELKGIAPAFRLSMFTFGD 1386
Db 1386 KLOSGVNLKEAFISAAETFFSWYYSKDGEGNELDKELKGIAPAFRLSMFTFGD 1376
Qy 1386 KVEYERIKREAFIOCSAETFLWKYKEDNNNGGEDLOJLESQIIPDFKROFEYFGD 1364
Db 1376 KVEYERIKREAFIOCSAETFLWKYKEDNNNGGEDLOJLESQIIPDFKROFEYFGD 1354
Qy 1364 YRDLFTGDIKSGHESKLEKQIDSLFKKNDQSPNCKTQOWMTESHEIWEAMCAL 1446
Db 1446 YRDLFTGDIKSGHESKLEKQIDSLFKKNDQSPNCKTQOWMTESHEIWEAMCAL 1436
Qy 1446 YRDLCTGDISSVG---NVQKINSALKRIGQSDDE---KKNMNGIKEDVWAGMVGCL 1419
Db 1419 YRDLCTGDISSVG---NVQKINSALKRIGQSDDE---KKNMNGIKEDVWAGMVGCL 1409
Qy 1419 SHDVGNNHKEIVRKIMEDPNNKXYNSVKTDEPGSTKISEAKYPOFLARLWTEWYD 1479
Db 1479 SHDVGNNHKEIVRKIMEDPNNKXYNSVKTDEPGSTKISEAKYPOFLARLWTEWYD 1469
Qy 1479 YCYTROYKLYOYCKSNQJLKDTECNKCEIYVYMK-KKEMJPODKYTYDERDKK 1552
Db 1552 YCYTROYKLYOYCKSNQJLKDTECNKCEIYVYMK-KKEMJPODKYTYDERDKK 1542

```

```

Db 1480 YCHTROYKLYEVESTKSDOLKCDTECNKCEIYVYMK-KKEMJPODKYTYDERDKK 1539
Qy 1539 REDROHIGVAVTDYTGNTADYLNKRTASCGDKPGSASVONRLOLEKQAYYADKHC 1612
Db 1540 RFNRHIGVAVTDYTGNTADYLNKRTASCGDKPGSASVONRLOLEKQAYYADKHC 1599
Qy 1613 GCTKFIENDDKTNTSSKCKGIVYEANTGAIKQONKPNYNNLKEJTEVLEPPSRRL 1672
Db 1600 GCTKFIENDDKTNTSSKCKGIVYEANTGAIKQONKPNYNNLKEJTEVLEPPSRRL 1659
Qy 1673 RICEFHALDGNVDPYKENGELKRLMEVAAEGYNLGOYKEKKEKKEKITSQAHKSY 1732
Db 1660 RICEFHALDGNVDPYKENGELKRLMEVAAEGYNLGOYKEKKEKKEKITSQAHKSY 1719
Qy 1733 EYPPGSAKSYDYLDIILGIDNLEDEKOTENLKIIFNKNGTSYVSGSDSTGNPGS 1792
Db 1720 EYPPGSAKSYDYLDIILGIDNLEDEKOTENLKIIFNKNGTSYVSGSDSTGNPGS 1779
Qy 1793 TARKFEMNENKCYVNNAMICYKRGDRDNGSNGSARSDDLKKGVSVDYDPYKGRND 1852
Db 1780 TARKFEMNENKCYVNNAMICYKRGDRDNGSNGSARSDDLKKGVSVDYDPYKGRND 1828
Qy 1853 EGTAYOFLRMEPEWGEDCKHKEKELVACNDYTGDNEDKRRKCTDACTOYKPFIS 1912
Db 1829 EGTAYOFLRMEPEWGEDCKHKEKELVACNDYTGDNEDKRRKCTDACTOYKPFIS 1888
Qy 1913 EKKPOYKOIKKYGNKRTIYSEHVAADAREYLDOLKICEKNSGDEYKCMKDV 1972
Db 1889 EKKPOYKOIKKYGNKRTIYSEHVAADAREYLDOLKICEKNSGDEYKCMKDV 1948
Qy 1973 STORLIDNSONMPASLDDEKEVEKNCOCVPPRGPRVRETPSPRSLSKTATKKE 2032
Db 1949 STORLIDNSONMPASLDDEKEVEKNCOCVPPRGPRVRETPSPRSLSKTATKKE 2008
Qy 2033 AKTAPPTOPKPVENLTTWAR--AOTRRAAQOTR-----RTSTATTESDVTMV 2083
Db 2009 AKTAPPTOPKPVENLTTWAR--AOTRRAAQOTR-----RTSTATTESDVTMV 2068
Qy 2083 KALISNKPDSRGIEGCPKTYGOYPRMGCIYVSKENENICMPRRKLCINNOYLN 2143
Db 2069 KALISNKPDSRGIEGCPKTYGOYPRMGCIYVSKENENICMPRRKLCINNOYLN 2125
Qy 2143 YETENKRDNDIKAFIFICALIETOFMLKYLIEENPAENELONGTIDPEKRIKYYTGD 2203
Db 2126 YETENKRDNDIKAFIFICALIETOFMLKYLIEENPAENELONGTIDPEKRIKYYTGD 2185
Qy 2203 YKDNFPGDIDISNDKIIITVNSVTIL--NENKCKQDK--KDELRKIFMEKKKFIWEG 2261
Db 2185 YKDNFPGDIDISNDKIIITVNSVTIL--NENKCKQDK--KDELRKIFMEKKKFIWEG 2245
Qy 2261 MYGLVTHLNDENKERTIRNOYQNDMTKLPSLEEVKRPOLRMTTEWAEFECKRRE 2321
Db 2246 MYGLVTHLNDENKERTIRNOYQNDMTKLPSLEEVKRPOLRMTTEWAEFECKRRE 2305
Qy 2321 QLKLEAGCEYEBCNGSNDKTOEACAVTYONFIKMWTEYEROREKPKKDKKRYK 2381
Db 2306 QLKLEAGCEYEBCNGSNDKTOEACAVTYONFIKMWTEYEROREKPKKDKKRYK 2263
Qy 2381 QLEMLKKECPDYTC---SVDSKTHECEDSCAKYOTFIKOWTYEKORGRKKNDDKDYK 2441
Db 2363 QLEMLKKECPDYTC---SVDSKTHECEDSCAKYOTFIKOWTYEKORGRKKNDDKDYK 2422
Qy 2441 LDYVPEEF--KCBPELSKSGSMITHKITEPKIIPANVCYKAAVYYSKEANNMOTLEA 2500
Db 2423 LDYVPEEF--KCBPELSKSGSMITHKITEPKIIPANVCYKAAVYYSKEANNMOTLEA 2471
Qy 2500 KFIPIESTKEKESKNWTN--NMPCDPKPYADKTIIGRRNPEENENRFRVDEYEMKY 2558
Db 2471 KFIPIESTKEKESKNWTN--NMPCDPKPYADKTIIGRRNPEENENRFRVDEYEMKY 2524
Qy 2558 KNSKFOYKRRVCPVPRRHEMLRNDIEKIRKDSYNTLLMVRRTANNEIDIIKNN 2618
Db 2524 KNSKFOYKRRVCPVPRRHEMLRNDIEKIRKDSYNTLLMVRRTANNEIDIIKNN 2594
Qy 2618 KERTNVDVTENICIIPRRKRYMCINPLENLGAKK--STLDLIFKRLKTAAYBSGHIKDSWD 2583
Db 2583 KERTNVDVTENICIIPRRKRYMCINPLENLGAKK--STLDLIFKRLKTAAYBSGHIKDSWD 2553

```

```

OY 2619 -SENCAMNPICDTMKYSPADIGDIIVGTDLR--IGYLPVEVILKYKVEFYIKWMN 2675
Db 2584 YKRNPKRRNOQLCDAMKYSFADLADIIIGTDIYKPIGA--NEVERKRALIFENIKYKMT 2641
OY 2676 KKKGNKRYVDVOTFSSAMADANRKOIMKAMTCAPADAKLFKGRMDGERTTLIODKCG 2735
Db 2642 ENKGIKXKHYDIOFPRAMADANREYKAMTCDAPDSMLFKLENGELIPNLTLSQHKC 2701
OY 2736 HADDPPVDYDIQFRRMAMTESEYCKALMELEKFKKSCDCKTSD-RCKNDYBENKCE 2794
Db 2702 YNDHPVDDYIQRRLMKKEGEYCKILNEKYNDKNDCKCKLMDKCSNDDEIKOR 2761
OY 2795 OCKTRCOEYKNFVLMKMSLFDIQSNKKYKELDYQYTKISTY-----DHVONFYOK 2845
Db 2762 SCKEKCKREYTKLIHMLKSGENIQKKKYNEL-----YTKIQNNRRGFINDNDKNVIEFEK 2816
OY 2846 LKTFSESEGVSESEYKLTSTCKLAKYKNEENDGSSNIRIYAEERPKSKKACSCILPSK 2905
Db 2817 VKMI-NNCVNGTPDXYLDKAIHCIAHNFQNGIKS--KPYAANMHEKYSKSCSTI-TH 2872
OY 2906 NPLDNCPTDQNDGCKELDTFTGCSKNDYNDNLDNMNAVLYNSSDDNKGVLLIPRRRL 2965
Db 2873 HPLDPCPNKTAAYCKTTHINPCITKNDNLETTGTGVVDNKKDKKGVLLPPRRRL 2932
OY 2966 CRRPTAYNRYKRGDEILKRLKLLTSASQGLIGOKYKSEELCEBAMKYSYADYSDIK 3025
Db 2933 CYRKLTKGNRYRINEKDNKLNLDLSAFSHGMLGKTFNDYNOGMSMKYSFADYADIK 2992
OY 3026 GIDMADT-----LSEKIKIFETSN-----EATENKRTWENNRRQIWMALCGY-KIAT 3075
Db 2993 GIDMIGGSNIDDFNKDKMPENNSENIGKTTISREOWMEERKKHVNAMLCGYOKGRK 3052
OY 3076 SVYTLDEGQCQPKDEETNOFLKMLIEMAKQACEKKHYSDSLKTCPSRNEDEFEASEL 3135
Db 3053 NNGEMDKNKNVPTEDGTQFLRMLIEMAMQAKYKNHVRDSIKTCRCSEKEDNFEASEL 3112
OY 3136 LKPGCGQNDIRYISILNLIKNTMENLITKYOLKQSSGNDINKRSEENVOYIYSKSDS 3195
Db 3113 LKPGCGQNDIRYISILNLIKNSMENLITKYOLKQPSGNDINKRSEENVOYIYSKSDS 3172
OY 3196 OCALENDINELVTGKNNENNEFKKLYRGLYFVDETHKKNVLDGNTKEEQOTR 3255
Db 3173 ECDLENDINELDTGKNNENNEFK-VLKKLYRGLYFVDETHKKNVLDGNTKEEQOTR 3231
OY 3256 PKALYFETPHVDSFOAPLESTHRVAOYDPKNDILKSSISVIVASLGLIALHEFM 3311
Db 3232 PKALYFETPHVDSFOARLFLPTRYREKYPKNDILKSSISVIVASLGLIALHEFM 3287

```

```

Query Match 38.0%; Score 7367; DB 5; Length 1615;
Best Local Similarity 83.0%; Pred. No. 0;
Matches 1358; Conservative 69; Mismatches 169; Indels 40; Gaps 12;

OY 12 MGNAAASLEGDAKSPILKSHKSARVNLERYAKNIRHPSKYAKEHVDLSKGLDTKAEFRG 71
Db 1 MNTESLSLEGEARSPSILIESSENSPRVNLERYAKNIRHPSKYAKEHVDLSKGLDTKAEFRG 60
OY 72 GSTPVRKNKNNYYPYPCNDLHKHTNLRYDYNVLRHPCGRGRGRNDREDEESCGKIKIN 131
Db 61 GSTPVRKNKNNYYPYPCNDLHKHTNLRYDYNVLRHPCGRGRGRNDREDEESCGKIKIN 120
OY 132 YKRNKADICAPRRRRHMDCKNLALNDINTQNIHDLGNVLTARYESIYNNPHRG 191
Db 121 YKRENDAICAPRRRRHMDCKNLALNDINTQNIHDLGNVLTARYESIYNNPHRG 180
OY 192 TSDACTALARSFADIGDIYRGIDMFKPNYHDKYETGLREVFKKINDGMEDEKANDYNDPG 251
Db 181 TSDVCTALARSFADIGDIYRGIDMFKPNYHDKYETGLREVFKKINDGMEDEKANDYNDPG 240
OY 252 SGNYKLRBAMVNVNKNKYWEAITCDASYKSGYFMOSESNTPLFSNPKCGHKQKYP-TN 310
Db 241 SGNYKLRBAMVNVNKNKYWEAITCDALPKRSAYFMOSEDKOLFSTPKCGHNNKDDPLTN 300
OY 311 LDYVPQYLKMFDEWGEFRCRRKRIKLKYKDSGRNDRERYLCSHNGDCTTITMKRGLIL 370
Db 301 LDYVPQYLKMFDEWGEFRCRRKRIKLKYKDSGRNDRERYLCSHNGDCTTITMKRGLIL 360
OY 371 LDKKCTDCTCKCVFVWMLGNQOBAFRRKOKERYEKEIYSTLSDNKFVNNINSEYKQFY 430
Db 361 LDKKCTDCTCKCVFVWMLGNQOBAFRRKOKERYEKEIYSTLSDNKFVNNINSEYKQFY 420
OY 431 EKLKETOYATNPDELNLNEGKCRGGLPGERKOTFTNSADDDGIRYSEYCOVCPDCGY 490
Db 421 DQLRDKNYKNDLDTFLNLBEGKCRKELKLGENDINTNSDDKGTTRYRQGYCOVCPDCGY 480
OY 491 KCDGIRYKHSNDREVRNNEDEKPPMGVYKPTNITVLYSGNEGODITOKLENCNSTNY 550
Db 481 KRGDTYTRKHLNDRBCVNNNEDKPPMGVYKPTNITVLYSGNEGODITOKLENCNSTNY 540
OY 551 KDKNNKMECYKYDENINCKLEONTNINNDPKIISFNHFLWYTYLLRDTIKANDKL 610
Db 541 KDKNNKMECYKYDENINCKLEONTNINNDPKIISFNHFLWYTYLLRDTIKANDKL 600
OY 611 KTCINNTTHCIDEGRNCLCPDRMYKOKEEENNSIKLFTKKNVPOPYTININLFBG 670
Db 601 KTCINNTTHCIDEGRNCLCPDRMYKOKEEENNSIKLFTKKNVPOPYTININLFBG 660
OY 671 YFEKVMKLDKDEAKWKEIMENIKRRKNFESNLNNRQVLENAIELDLHKLKETAATCKD 730
Db 661 YFEKVMKLDKDEAKWKEIMENIKRRKNFESNLNNRQVLENAIELDLHKLKETAATCKD 720
OY 731 NNTNEACETSHNATTPCYKPRGCTOPTKNIKIAOYFKRSAYEERANGLHKLKKAHE 790
Db 721 NNTNEAYETSHNATTPCYKPRGCTOPTKNIKIAOYFKRSAYEERANGLHKLKKAHE 780
OY 791 GIYKRGRRRDFDNCRCRIMIKSNRNLFSPNCPGCKGKGDOIGORFYVGTMEVDEPHE 850
Db 781 GIYKRGRRRDFDNCRCRIMIKSNRNLFSPNCPGCKGKGDOIGORFYVGTMEVDEPHE 840
OY 851 MKRDHEDVIMPPRRRHICSTNLEHLOTDPHPLNGNITVDLLVNNSPFGDVLISAKYEAANKI 910
Db 841 MKRDHEDVIMPPRRRHICSTNLEHLOTDPHPLNGNITVDLLVNNSPFGDVLISAKYEAANKI 900
OY 911 IRNYKRNKNNLKGPEYTDPRKHQTTICRAIRYSFADIGDIIRGDLWERNGDWYKLGCHLE 970
Db 901 IRNYKRNKNNLKGPEYTDPRKHQTTICRAIRYSFADIGDIIRGDLWERNGDWYKLGCHLE 960
OY 971 TYVGNITKSLKGGNDYNDADAKYLLKLENNWEANRAKAYEAMKCDIYLLKDKSGHST 1030
Db 961 TYVGNITKSLKGGNDYNDADAKYLLKLENNWEANRAKAYEAMKCDIYLLKDKSGHST 1020

```

```

QY 1031 QSSYGYSDHPLDLYIQKLRMTAEAWCYKQYKKEKCEKCKDKNDGQCTK 1090
DB 1021 QSSYGYSDHPLDLYIQKLRMTAEAWCYKQYKKEKCEKCKDKNDGQCTK 1080
QY 1091 ESGTCTCTACNEYNIIIGLTKRMWNIISDKYELHEOQMSVNSGIEASSAKNHI 1150
DB 1081 ESGTCTCTACNEYNIIIGLTKRMWNIISDKYELHEOQMSVNSGIEASSAKNHI 1140
QY 1151 DRNVEFLSELVQONGKSNKSGTSDSASVIGTNTYENVAGYLLHDGTFDQCSQNEFC 1210
DB 1141 DRNVEFLSELVQONGKSNKSGTSDSASVIGTNTYENVAGYLLHDGTFDQCSQNEFC 1200
QY 1211 DEKSGKDNKEXAFRDKQDDHAGCGSSGSKPTVQJTKTKKAEKTECKTVDDIIE 1270
DB 1201 DEKSGKDNKEXAFRDKQDDHAGCGSSGSKPTVQJTKTKKAEKTECKTVDDIIE 1253
QY 1271 NDGRKOVEDCHPRKNSNGYPRMOC---GNINIVEDPRVACAPPRRKLCHVFLANDNEIK 1326
DB 1254 NDGRKOVEDCHPRKNSNGYPRMOC---GNINIVEDPRVACAPPRRKLCHVFLANDNEIK 1304
QY 1327 KLSQOVNLKEAFIKSAAEFEFSWYYSKDEGNEIDKEKGIIPAPLRSMFEYFGD 1386
DB 1305 KVEDEKREAFIOCSAAEFELMKKYKEDNNGGEDLQNLSEGIIPDPFRKMFYTFGD 1364
QY 1387 YRDLFLGTDIKSGHSGSKLKEOJDSLFKNDOKSPNKCTQEWTEHSHETWEMALCAL 1446
DB 1365 YRDLFLGTDIKSGHSGSKLKEOJDSLFKNDOKSPNKCTQEWTEHSHETWEMALCAL 1419
QY 1447 -----VKIGAKKDDPTENGYNNVAFSDK--STTEEFARPOFLRWLEWDD 1493
DB 1420 SHDVSQNHKEIVRKKIMEDPQNNKYQVNSVAFTEDEPSTKISEFAKVQFLRWLEWDD 1479
QY 1494 YCTRYOKYLAKVOEKCSNDOLKCDTECNKKECEDYVYKYM-KKKKEWIPQDKYVYDERDK 1552
DB 1480 YCHTROKYLAKVESTCKSNCDLCKDTECNKCEDEYKMKKKKKEMWILQDXYKDERDK 1539
QY 1553 REDSOHIGVWYTDYTNATNYLNRKFLPASGCDKPGASVYORNIQLEKRAYYDADKHC 1612
DB 1540 RFNOHIGVWYTDYTNATNYLNRKFLPASGCDKPGSSVYORNIQLEKRAYYDADKHC 1599
QY 1613 GCTKEIENDKYNINIS 1628
DB 1600 GCTKEIENDKYNINIS 1615

RESULT 4
O15870 PRELIMINARY; PRT; 2706 AA.
O15870;
01-JAN-1998 (Tremblrel. 05, Created)
01-JAN-1998 (Tremblrel. 05, Last sequence update)
01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PfEMP1 (Fragment).
OS R29R+VARL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=IT 4/25/5;
RX MEDLINE=97373957; Pubmed=9230440;
RA Rowe J.A., Moulds J.M., Newbold C.I., Miller L.H.;
RT "Plasmodium falciparum" rosetting is mediated by PfEMP1 and requires
RT complement receptor 1."
RL Nature 388:292-295(1997).
DR EMBL: Y13402; CAA73831.1; -.
DR EMBL: Y13403; CAA73831.1; JOINED.
DR InterPro: IPR002086; Aldehyde-dehydr.
DR InterPro: IPR004258; PfEMP.
DR Pfam: PF03011; PfEMP; 2.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
FT NON TER 2706 2706
SQ SEQUENCE 2706 AA; 308162 MW; 1C3D55AD5317D68 CRC64;

```

```

Query Match 23.3%; Score 4515.5; DB 5; Length 2706;
Best Local Similarity 31.8%; Pred. No. 4,4e-210;
Matches 1191; Conservative 372; Mismatches 871; Indels 1315; Gaps 123;

QY 34 SARNVLEKXAKNIRPSK---YAKEHVDLSKGLITKAEF-----RGSPSTPVNK 79
DB 14 SATDVEKIAFGYNOEKEKYPEE--NELKILSNALFVQLRKELNIESPSPSD----- 67
QY 80 HNYTTPYCNLDHEKHTNLARDDVNLRHPCGRGRQNRDEDEESBCC- NKRIRNKRKDA 138
DB 68 -----SCSDHNFHTNINTETEGRPCKYERNEKRSNGEAKCGSDKLRDGIISAG 120
QY 139 IACAPRRRHHCDNLNLEALNDINTONIHDLGNLVAKYEGESTLVNHPKGT-----SD 194
DB 121 GACAPFRQNLCDNRLEYLELKNKNTHTDILGNLVAKYEGSDLVNHPKNSGNKSS 180
QY 195 ACTALARFAIDIGIVRGIDMFKRVNHDVETGLREYFKKTHDGNEDVYKDY- NPDGSG 253
DB 181 ICTALARSFADIGIVRGIDMFKRVNHDVETGLREYFKKTHDGNEDVYKDY- NPDGSG 240
QY 254 NYUKLEAMVNVNRRKRVWEATLTCDAYSKYGFQMSSESTPLF-SNPKGHQGVPTNMD 312
DB 241 NYUKLEAMVNVNRRKRVWEATLTCDAYSKYGFQMSSESTPLF-SNPKGHQGVPTNMD 300
QY 313 YVPOYLRFWDEMGEEFCRKNIKLKVKYDCSNDKERLYCSHNGHDTTITMKKGIHL 372
DB 301 YVPOFLRFWDEMAEFCRIRIKILENYKKECRDPENNKCYGCGDHCDKRYLKNITFID 360
QY 373 NKCDSCSKCYVFEVWVGNOGFAKKQKEKEKELQSTLSNDKFNVTI-NSEYKQFYE 431
DB 361 LNCPRCENACSNYTKWEIQRKQFQKQKRYKMEI-----KIKTINSNENDKEEYE 412
QY 432 KIKETQYATNTPTNLNLEGYCKGLPGEKIDFTNSADKGIIFYSEYQGVOPDGVYK 491
DB 413 NLDKKGISTITPLESLNKGOCODNIDKKNKTFKNLE--TFGSGYCEACPITYGVK 469
QY 492 CDGIKYTHKSDNDRERVNEDYKPPW-----GVKPTNITVLYSGNEOGDITOKLENCNS 546
DB 470 CSNEKCTPVTENEW-----NSNRRLEPTDTSFKNLATNIDMLVNGIGALINLEKKNCTK 525
QY 547 STNKKDKNNQKWEYCYDENINRCKLEONTEINN---DNPKIIFSHNFEFLMYTLLRDT 603
DB 526 YGLIKGIRKKQWQCOYLN-NIDCKI--NNVYMNNGYFDNK--IAFNVLFOHMYEYVRDH 580
QY 604 IKWMDKLKTCI---NNTTHCIDECNRRCLCFDRVWVOKKEEYNSIKKLFKKKNIO--- 657
DB 581 NRLKEKIDVCIKKENINENICIRKCTNCEGVGKMLEKKEAMDKINQHYNOKNHIIFIL 640
QY 658 -----QSTYSNINNLFGSYFFKVDKLDKDEAKKELMENIKRKNKESNLENNRYLEN 712
DB 641 IPWITGFYEKIT--FPNDFKALEVDVTIN----- 669
QY 713 AIELDLHLEKATATICDNNTNEACETSHNATNIPCVKPRGGTPTKNIKEIAQYFRSA 772
DB 670 -----VLDPLKE-----CQDTH----- 681
QY 773 YEARNRGLHLKGAHEGIYKRGRRKDKDNLCRITMIKHSNRLGFSNGPCDGKGTGD 832
DB 682 -----CKI----- 684
QY 833 GIOTRFVYGVBEVDPHMKRKHEDVIMPPRRRHICTSNLEHLOTDDHPLNGNIVDVLN 892
DB 685 -----EKIRSIDVLI----- 695
QY 893 NSFGLVLLSAKYEANKIIMYKKEKNLKPKEVTDPKHOTITICRAIRYFADIGIIRG 952
DB 696 -----KEIIS----- 700
QY 953 RDLERNGDWVQLQGHLETFVGNLHSLKCGKNDKYDADAPKYLKLENNWEANRAKWE 1012
DB 701 -----WLOKRIEV-- 708

```


0Y 1013 AMKDOI KYLKD KSGHOSGYSVGDHPTPLDYPQKLRMTENAEWCKYQKREYDKL 1072
 Db 709 -C-----KSHHEDDHEYC-----CDLIPKSVDDO 732
 0Y 1073 KEKCEKCKDKNDQOGCTKESGCTKCTEACNEYNIDILMKEQWNIISDKYKELEHQAO 1132
 Db 733 EEDDEY-----DEEKE-----744
 0Y 1133 MSVNSGIEASSTAKNHDNRNVIIEPLSELYOONGKSNKSGTSDSAVIGTNTYENGA 1192
 Db 745 -----SSQTTKRNI-----SOKGTFKASACVACAIYGV-L 777
 0Y 1193 YLHDNGNPDCCOSQNEFCCEKSDGKNEKAFRDKQDDHDGACGCKSGSKPTVQJKTAK 1252
 Db 778 OOKSNGSINDCNKAKN-----KN-----EMQCDKNTFVDSGCVCPRR 793
 0Y 1253 KAEKOTECKTVNDILKENDGKQVEDCHPKKNSNGYPMOCGNILVE-DPRVCPRR 1311
 Db 794 -----KN-----EMQCDKNTFVDSGCVCPRR 818
 1312 OKLCVHFLANDNEIKKLOSOVNLKFAIKSAAETFPFSYKSKDEGNEIDKELEK 1371
 Db 819 KSICINHLLEBQTK---NKYOLREAFIKCAKETMLMDKYKNDKNEAEEL---LKKG 872
 0Y 1372 IPPAFKSFYTFYTDYRDLFTGIDISKHGBGSKLEQJDSLFKNGDQKSPNGKTQEW 1431
 Db 873 IPDEFKRIEYFEGDEPDLCELENDMGK---DVKYKKNINKKFNNSKRGFKKIDPENW 929
 0Y 1432 TEHSHIWEAMCALYKICAK-----KDDPTENYGYNNYKFSDK-----STLLEFAKPR 1481
 Db 930 NENGPRIHNGMICALIHADTKDSIRKND---NYKTEKTYILAKRGSGNMTLSEAKPR 985
 0Y 1482 QELRLMTEYDYCYTRQYKLYADVOEKNSD---OLKCDTECNKCKEDYKTKKKR-EM 1538
 Db 986 KFLMFEVEMYDYCKEROKYELFEVASTCKSIDGQLKCDRGCKNNKDEYKYYRKKKEE 1045
 0Y 1539 IPQDKYXDERDKRRDRHIGVMTDYGTNAITDYLNRKFTASCDCDKRGA-----SVY 1593
 Db 1046 NIQDKYKXDERDKRRDRHIGVMTDYGTNAITDYLNRKFTASCDCDKRGA-----SVY 1593
 0Y 1594 ORNIQLEKQAYYADADKHCCTEFTENDKTYNISCKDKCKGLVKANTGAIWQKGN 1653
 Db 1106 KNIIELESEGYADQYGCCTFI-HDKYKTSIGRSNCCGLNSAKKNIKWRSDER 1164
 0Y 1654 NYNNLKE--LTEDVLEPPSRRLICFHALDGNVTPDYKDNGLRKLRLMEVAATEGYNLQ 1711
 Db 1165 DVAFLKRLMSGVPFSRRLRICFHALDGNVTPDYKDNGLRKLRLMEVAATEGYNLQ 1224
 1712 YKKEKKEKIKTSOAHKYSYEPYPSAMKYSFYDLROTLIGDNLDEKQKTEEMUKI 1771
 Db 1225 YKKEKKEKTEATEEAHAKYSYEQVPCSAMKYSFYDLROTLIGDNLDEKQKTEEMUKI 1284
 0Y 1772 FNRKGTSGVKGSDSTGNGSTARKFFNMENKCYVNNAMICGYKRGDRDNGSNGSARSD 1831
 Db 1285 F-----KSESNGSNGSNGSTARKFFNMENKCYVNNAMICGYKRGDRDNGSNGSARSD 1332
 0Y 1832 DLKKGVSVDSDYPMGNKRNDEGTAYQFLRWEFWEWEDFCFKHKELEKLVGACNDYTCG 1891
 Db 1333 DLKKGVSVDSDYPMGNKRNDEGTAYQFLRWEFWEWEDFCFKHKELEKLVGACNDYTCG 1392
 0Y 1892 DNEEDKRCCTDCTOYKFFITSEMKPOYEROKIKYGENKIKYSEHVAADADEARAYTLK 1951
 Db 1393 DNEEDKRCCTDCTOYKFFITSEMKPOYEROKIKYGENKIKYSEHVAADADEARAYTLK 1452
 0Y 1952 OLKKEICENSGDEYKCMKDVSTQRLTDSNOSMPPASLDEPEKEVGKCNQVPRPRV 2011
 Db 1453 QLOKSC-NSGKCD-CMKNKSF-----SNGNMNPASLDEPEKEVGKCNQVPRPRV 1502
 0Y 2012 RRTTSPRSLSKATASKKEAT-----APPTKQPK-----2044
 Db 1503 ---PAPAPRPARRESGNVGRSEGEDGPLPLPPLPPKPGGAGAGRILSTPRNG 1559
 0Y 2045 --VENLTERAQTTRBRAAQOTRKRRTSTATTESVGVNVAKILSNKPDPSRGI-EGCN 2101

Db 1560 TIEDEEDDEGEKAEALASEETSK--EVEQKEDTTEKCYKIVANILTGKONLDACN 1617
 0Y 2102 PKTYGQYPR---WGCIV-----GSKENENNGICMPRRKRLK 2135
 Db 1618 QK-YG-YQPRHMKWCVYTTSSSTSERGASNNKNDSTKSDKNGSICIPRRKRLX 1675
 0Y 2136 INNIO-YLNYTE-----NKRNDIK-----EAFIKALETOPLML 2171
 Db 1676 IKKIQEMASGNQACNGTSGDSTGASNPQNGVSTSPQVALLHAFVESAAYETFLMD 1735
 0Y 2172 KY-----IENPAE-----NELQGTIPDEKRIIMYTYGDKYKMFNG 2210
 Db 1736 RYKKEKEIKQOQETGLVASETSEDTEHPQNKQJGTLPLDLRQMFYTLGDRICVG 1795
 0Y 2211 TDISNKKIITVNSYTTLLENKKKOD--KKD-----EELKLFEMKKNFIEGMI 2263
 Db 1796 ---KTPDGIDITYSASDKDMDKIOAKIOQLPKKIDPPSVKPTPOQYWNKHAESINNGMI 1852
 0Y 2264 YGLTYHLTDENEKIR-----DNYQNDMTKL-----2291
 Db 1853 YALTYR-TTPSGEKPKQIPEVTKLFEDEKGTPOSNKYQYKV-KLEEBETSGAKPKSTE 1910
 0Y 2292 --TPS---LEEFYKRPQFLRFTMAEBEFCCKRKEQLKLEAGC---KEYEONG-- 2337
 Db 1911 SSSPSGEMTPLDIFISRPFRFLYEMGQNFCEKRRKRLGDINCLKDGDKQYSGDGA 1970
 0Y 2338 -SN-----DGKTOEAEACVYTONFIKMKTEYEROREKPKKDKGKKY--DY 2383
 Db 1971 GSNIDVNMKIDFIDLGPRCAKPCSSYRKWKIKKQYERQOEAAYQOECCKEYNNH 2030
 0Y 2384 PSTERDIEKATCAHEYLNMKLELGNKDCSCQKPSOLPKTTQOSSDANDMPESID 2443
 Db 2031 NGFRVILGCTTAPGDL---QTLKNGPC-----KSENGRDEE-- 2065
 0Y 2444 YPBEENKCEPBLSKSGSIHHTKITEPIPNVCYKAYISKRAENMDITLKEPI 2503
 Db 2066 --DELNFSO-PDYTRPA---TNCPTCPKFKYNC-----2093
 0Y 2504 PIESTKEESKNSMTNNPCDPKPAAPKYIGRRNPCEENREPRKVDYEMKCYKNSKF 2563
 Db 2094 -----KNGNCADJNGKNGKTP-----2111
 0Y 2564 YOEKKNVCYPPRREHMLNLEDEIKTERLKDSTYLLKMYRRARNEGDITIKNFSENGC 2623
 Db 2112 -----IDAQNEOMQOTAKEFVMIYVSDKSTNGFEV---NDLNEC 2147
 0Y 2624 AMNPICDTMKYSFADJGDIYRGDMLRIGYLPPEVLEIKLYKVEEYLYGKRNKNGRNY 2683
 Db 2148 A-----GAD-----IFGI-----2156
 0Y 2684 NDVQTPRSAMWDANRRKDIWKAM-----TCKADE---DAKLFKRGMDGFERITLIQDKC 2734
 Db 2157 -----KENINSGRNVGLDYCK-PEKYNNDQKYNK-ENDG-----2189
 0Y 2735 GHKDDPPVDYIPQFRMWTMESEYKAKAMEELEKFKKSDCHKTSDRCKNDYDENKCE 2794
 Db 2190 -----YIIQIALLRNVENF---LEDYKKIKRHSICHTNSTEEK---KSTCD 2232
 0Y 2795 QCKTRCOEYKNFLKMSLEJDIOSNKYKELYEDPIYTKISTYHVNQNFVQKLTFFSECS 2854
 Db 2233 -CGKCKC---CGQWIKLKEEEMKIKENY-----VEKNSE-- 2265
 0Y 2855 VESFSEYLETSKLNYKFNENDSSNIRTYAEETPKSYKACSCGLPSKNLDCNPTD 2914
 Db 2266 -DQFSDNLNLFLETLITQIPVADVQGVYIKLSNFDPP-----CGSADANS-----2310
 0Y 2915 QNKDQCKELOTFFGCKNDYDNNLDMNAYLVYNSDDNKGCVLIPRRRLCTRPITYAN 2974
 Db 2311 QKRDG-----NENAIQDCMIN-----RLOOKAKNCTOPSGSKOCTTPTPTL- 2352
 0Y 2975 YKRGDEILLKKLJLTSFSGQLGQKYKSEELCEFAKMYSTADYSDIIGTDMDJSL 3034

Db 2353 -----EDDETFD-----DDI----- 2362
QY 3035 SEKIKKIFETSNATEENRKTWMENNRQIWHAMLCGYKATSAVTLDEGCOLPKDEETN 3094
Db 2363 -----ETDN-----PVANPICGSIOTOTETVLEECVAPRAPEEK 2399
QY 3095 QFRLWLEMAKQCKEKKHVSLSLTKCPRSNEDNEASELLRQCONDIRKYLISLNL 3154
Db 2400 D-----EKKREDP----- 2408
QY 3155 IKNTMENLNKIKYKOLDOSSGNIDN-----PSEENVQSYIKSKDSOCALENDINEIVTG 3210
Db 2409 -----AEBOGAIQSPGAPGPPSPPTS----- 2431
QY 3211 TKNNENNEKVLKLYLGLYVEDETHKNHVLGNIKEEQTVRKALYFTPHVDSFY 3270
Db 2432 -----DOTENHV-----TPREDDPP 2447
QY 3271 QAPLSTHVAQ-----YDKNDILKSSIVYVSAALGILALHEKKKFKSSVLL 3321
Db 2448 PAD--DTRPPSPPLPPADQPPDP--TIIQYIIPGVALALGSIATFAFFMKKKTKHPVDLF 2503
QY 3322 RLINIPQEGEMPTLESKNRYIPYRSGPYKGYIYMEGDTSGDEKXVWMDLSSSDITSS 3381
Db 2504 SVINTPKGYDIPTLKSKNRITPYRSGSYKGTIYMEDS--DEDKTAFMSDITDITSS 2561
QY 3382 ESEYEEILDINDIYVPGSPYKTLIEVLEPSKR--DIP--SDDTSPND--TPRTNRFIDEM 3437
Db 2562 ESEYEEILDINDIYVPGSPYKTLIEVLEPSKSNGINIPHAGEPLDDMVPPTNFTDEEM 2621
QY 3438 NELKHEFVSQYLPNTEPN--NYK--SADIEMTEPNTLXSDNPEKPFITSHDLYTG 3494
Db 2622 NELKHEFVSQYLPNTEPN--NYK--SADIEMTEPNTLXSDNPEKPFITSHDLYTG 3494
QY 3495 KEISTYNINM--VNNIDIPMSDKNGTYGTI 2706
Db 2681 EISTYNINM--VNNIDIPMSDKNGTYGTI 2706
RESULT 5
Q26032 PRELIMINARY; PRT; 3006 AA.
ID 026032
AC 026032
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
DE Variant-specific surface protein.
GN VAR-3.
Plasmodium falciparum.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5633;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.2., Heatwole V.M., Wertheimer S.P., Guinet F., Herfeldt J.A.,
Peterson D.S., Ravetch J.A., Wellem T.E.;
RT "The large diverse gene family var encodes proteins involved in
cytoadherence and antigenic variation of Plasmodium falciparum-
infected erythrocytes."
RL Cell 82:89-100(1995).
DR EMBL; LA0609; AAA75397.1; -
DR InterPro; IPR004258; PFEEM.
DR Pfam; PFO3011; PFEEM; 2.
SQ SEQUENCE 3006 AA; 343769 MW; 8PD5FP475F943C74 CRC64;
Query Match 21.9%; Score 4254; DB 5; Length 3006;
Best Local Similarity 30.7%; Pred. No. 2.3e-197;
Matches 1133; Conservative 483; Mismatches 1002; Indels 1076; Gaps 133;
QY 31 SHKARNVLERYAKNIRH-PSKYAKHEVDSLKGDLTAERGGSTFVNKNHNTYPPCN 89
Db 8 TNKTAKEVLDIGKEVOEKATEDALYTRNDLGNTLSQAKFHG--VPIDVKN-----PCD 59

QY 90 LDHKEHTNLRYDDVNLREPCHEGRQNRDEDESESCGNKIRNYKRRKNDALACAPRRRH 149
Db 60 LVEIHTNV-VGRKKENPCRGREERFSDVLSGQCAKNTIKDSVTNSGACAPRRRLH 118
QY 150 CDKNLEALNDINTONTIHDLGVLVTAKEGESYVNNPH-----KGTSDACTAALARSAD 205
Db 119 GDRNELIKTDENTSTHDLVDLTAHGBESLVKHKEYIKKRNRPNTCTVLARSAD 178
QY 206 ICDIYRGIDMF-----KPVHDKVETGLREYFKRIHDM-----EDEVKNDYN 248
Db 179 ICDIYRGIDLYGYDEKEERKOLEKMLKFFQGIHDVAKTSGRTNGKSAEQAQRYN 238
QY 249 PDGSGNYKLRBAMNVNPNKRYWEAITCDASYKSYFQSSNPPLNSPNCGRKQKVP 308
Db 239 -DATNGYTKLRBAMNVNPNKRYWEAITCDADNDEYFENSSDGLYVPSNGCGRNEKVP 297
QY 309 TNLDPVQYTLNMFDEWGEFCRKRNIKLKRYKDSGRNDEKELYSNGHSDCTTIWKGI 368
Db 298 TNLDPVQYTLNMFDEWGEFCRKRNIKLKRYKDSGRNDEKELYSNGHSDCTTIWKGI 368
QY 369 LHLNKKCTDCTCKAVEFWLGNQOEAQKQEKYKEIYOSYLSNDKFNVINSEYKQ 428
Db 358 CSSDNCCTACSNICLAYDAMLNORNEFEKOKIKYKEIEKYSKSSDKSNSINIKYNE 417
QY 429 FYEKLETOYATNDFLNLNEGKYCKGGLDEGDIFTNADKGIIFYSEYQVCPDC 488
Db 418 FYENGRKYEYTLQNLFLKLNKMGTCOEKIEEVEVDF--NKDDMVHFHSEYCOCPDC 475
QY 489 GVKCDGIKYTKSDNDRR-----VNNEDYKPPMGVKNPTNIVLYSGNEOGDITOKLE 541
Db 476 VVOCCKGCKTBDKNDKCRSKIKILQSE-----PLEIHVLSMDKOGDITOKLE 527
QY 542 NECSSTNYKDXNNQKWEYCKDENIRCKLEONTETLNNDPKTISHNPFELAVTYLRL 601
Db 528 VFGSSTNYBERNVOQKWKCNKNSDYNCCENISSYSDSTANVMSVECFHSMAKILLI 587
QY 602 DTKMNDKLKTCINNT--THICIDECNRCICFDEWVQKKEEENSIKTLTKKKNIQSY 660
Db 588 DTIKWEHQKNCINNTVTCESKICNCEYEMIKRHEHEKAYKANNRRNSYIT 647
QY 661 YSNINNFEGYEFKVDKLDKDE-AKKKEIMENIKRRKNEFSNLENNRDYLENAIELLD 719
Db 648 YNNISRYVDSFLQVMAALDODEGKWDPTEDL--KKFEPSTNPTPGSQDAIEFLD 706
QY 720 HKEFATICKDNNTNEACETSHNATNCPYKPRGGQPTNINIEIQYFRSAVEARNR 779
Db 707 HLDNALITCKDNNSNESCDSKVKTNPCGKNPSASNNLVYRRLAEMQRYARKOLEKR 766
QY 780 GLH-KLKGAKHEGTYKGRGRKDFKDLIRITKHS---NRNLGFS-NGPCDG--GTGPG 833
Db 767 GGEINLKGDATKGYRQGGPADGK--NVCISINQNHNTVQNNNAAYYQGFCTGKDSNG 825
QY 834 IOTRFYVGTMEVDPELMKRDHEDVIMPPRRHICTSNLEHLOTDPHL--NGNIYDVLN 892
Db 826 V--RMKGTPEWK--PGRIQMSADDIYMPRRQHMCSNLEYLOTGKPGKQD--GKLVN 880
QY 893 NSFGLVLSAKYEANKIIRMYKEKNNLKGPRKEVTPKQOTICRAIRVSPFDIGIING 952
Db 881 NSFGLVLSAKKADKADKIIELYKKNN---KSVLTDPEDNESACRLRYSFADLGIING 937
QY 953 RDLMEKNGMVKIQLHLETFVGNHRSKLGKGNKNDADAPYKLKLREWMENANRAKYWE 1012
Db 938 RDLMDKNSDAKRLQTLNKEIFTKIEELPEDIKKKYDKGTGTHKLRLREWMENANRQVWR 997
QY 1013 AMKCDIKYLLKDKSGHSTOSSYCGSDHTPLDITPQKIRNMTWNAEWYCKAQKYEYKL 1072
Db 998 AMKCALENDKDK-----CNGIPIDYIIPQRLRMTEVAEWCKQSDSLYNYL 1045
QY 1073 KECKECKDKNDGOGCTKESGTCTCTEACNEVNDIILAMEONNITSDKKEIHEQO 1132
Db 1046 VADCKSCCKK--AKSTQDNG--DCTFKCAACDNYNKKIKPMEQOEKIKNKTAOLYKAL 1102

QY 1133 MYSNSGIEASSTAKNHNIDRNVEFLSELYOQN-GGKSNKSGTSDSAVIGTITTYENVG 1191
 1103 DSYNGEESKSKKTAQADQVHFLAELTRKSGGKGNVKTIVSPTTPTTLVSSAA 1162
 QY 1192 AVL-HPTGNFDDQSONEFCEKSDGKDNKRYFRDPODHGACGKS-GSPTRVQIK 1249
 1163 GYHHELGKRVGNCNTOKEPYSK-----NKTAFFKDPKGYEACKCNDRNRP-----Q 1212
 QY 1250 TKRAAEDECTEYVDILKENDGKROVEDCHP--KNSNGYEDMOCGINLVEDEPRVCM 1307
 1213 PAKKDEDA-CDVAPPLAKDETDIDGCONKRYKAGKDKYGMCONSGIHTHNGACM 1271
 QY 1308 PPRKOLCHAFLANDNEIKKLOSOVMKEAFISAAEFPSYKSKDCEGNEIDKEL 1367
 1272 PPRKOLCVSGLTKTRIKAE---YIRTEFIKSAIETHFAMDRKEDNGEA---EAEI 1325
 QY 1368 KEKTIAPAFRSMFYEGVDRDLFGDLSKSGESKLEOJDSIF-KNGDOK-----S 1421
 1326 KGNITBEGFRKQMYTFGDRDLFGDLSKSGESKLEOJDSIF-KNGDOK-----S 1384
 Db 1422 PNGKROEMWTEHSEIWEAMCALV-KIG--AKKDFEENGYNNVYKESDK-STLLEEF 1477
 1385 SNNEILLDDMDQKIDMEGMLCALTHKISDEEKKKEIKNKYSYKRLNESPKGSNKYEDF 1444
 QY 1478 AKRPOFLRWLEFYDYCYTROYKYLKDYOEKC--KSD--OLKCDTECNKCEGYVYK 1531
 1445 AKRPOFLRWLEFYDYCYTROYKYLKDYOEKC--KSD--OLKCDTECNKCEGYVYK 1504
 QY 1532 MKKKKEWIPDKYKDERDKRFRDHOHIGVAVDYTG---TNAIDYLNKRFETASCD 1585
 1505 IYTKV---EYTK---QKGFDAEKI---IDKBYEBSFTKDSSEYIAKK---C-- 1546
 QY 1586 KPGSASVYORNIOLLEKQAYYADKHGCTYKTEIENDKYTNISSKDKGLVKEANTGAI 1645
 1547 -----LDDTCMCQKYNKNTYENT----- 1566
 QY 1646 KMNKGPNNNNNLKELTEVLFPSRLRICFHALDGYTDPYEVKENDGLKRLMEVAATE 1705
 1567 --BNKYTTN-SNLEKRE--CQPP-----EPPGPEGAR-----SDS 1600
 QY 1706 GYNLGOYKKEKKEKIKTSDAKHSEYEVPCSAKMYSPFDLDIILGIDNLEDEKOTE 1765
 1601 G-----PROTPRAGSDASNTVPSPPRAGDTHVAV-----QEEBEDE 1645
 QY 1766 ENLKJFNKNGTSGKSDSTGNPGSTAKRFFWNNKCECVNAMIYGRGRDNGSN 1825
 1646 -----DGLPE 1651
 1826 SARSDIEDLKCGSPSDDYPMGKNRDEGTAYQPLRFAEM--GEDPCKHKEKLEKLVGA 1884
 1652 D--ODEVEVAGAEEED-----LDVGA-RILGRNTSPDEDEEASEE----- 1694
 QY 1885 CNDYCGDNEDKRRKCTADQYKFKISEWKPOYKOIKRYGENKDKIYSEHPVAKDAED 1944
 1695 -----DDDDDDADPTTEVYG-----GEEETAEHDOD 1721
 QY 1945 AREYLDKOLKICEBNSGDEYKCMADVSTORLTDGNSOMPASLDEPEKEGKCNQY 2004
 1722 TTEETVQ--EKAEDKDGGE-----TPOKET-----QPEVEYNPCD-- 1757
 QY 2005 PRGPRVRRTSPRSLISKATASKREAKTAPTKOPKVENUTTEMROTRRRAAO 2064
 1758 -----YVTLFTTETLEACPTKVNGRE----- 1782
 QY 2065 TRKRTSTATTESDVGTMVAAILSNKPSDRGIEGCPKPYGOKPKGCIYVGSKENENG 2124
 1783 -----KFPNNKCI-SSGSDASGS 1799
 QY 2125 ICMPPRRKILCINNIOYLT-----ETENKRDND-KEAFICAALETOTFLWKY-IE 2176
 1800 ICIPPRRRKLYLHIEBVDTVSSDGETTPIHDALREAFIOTAAVEFELMHRYKIK 1859
 QY 2177 NPAENELONGT-----IDDEKRLIMYTYGYDKMDFG 2210

Db 1860 EKEROEELONGTLELPPAQKVSPEDNPEHOKKLEKGIKPEEKROMFYTLGDRLCVG 1919
 QY 2211 T--DI-----SNDKITVTNSVTTILNENN-----KKKOKKKDEELRKIE 2250
 1920 VKDDVAQALBASDKNKSGDKNIDISEKISYIEKSGEOTPPGPKGQITTKPEE---W 1975
 QY 2251 WEKRRPFMEGMYGLTYLTD-----ENEKIRDN-YOYNDMTRKLPSS-- 2294
 1976 WOKNGEIHMMAMICALT-HMTDFROYDDQYKGLGFENKGNTPANSQOYKRNVTISSVSN 2034
 QY 2295 -----LEEFYKPOFLRWLEFYDYCYTROYKYLKDYOEKC--KSD--OLKCDTECNKCEGYVYK 2341
 2035 GGPIGNIKLEOFASRPTFLRWLEFYDYCYTROYKYLKDYOEKC--KSD--OLKCDTECNKCEGYVYK 2094
 QY 2342 -----KTOEACVYTONFIKMKTEYEROR---EKFRKDKDK--YKOD 2383
 2095 EKPCKDGSFETLKCPSCASCSYKMWISRKDETFKQAGAYEKOKDABGNNDYKER 2154
 QY 2384 PSTERIEKATCAHEYLMMKLKELCGNKDCSCMOKPSSOLPKTTOOSSDANDMBESLD 2443
 2155 SKTLRNYNDAAA---FLN-SLK---NGPCS-----KNDSDVQDEIK 2189
 QY 2444 YVPEE---FNKCEPBLSKGSMIHTKITEPKIIPANCVEKAYILSKAENMDITLKE 2500
 2190 FDEKRTFGEHYCKPCKSK-----ITVCK 2213
 QY 2501 KFIPIESTEKESKNSVTNNPCDPKPYADKYIGRNPCEENREERFKVDEYEMKCYKN 2560
 2214 K-----ENHCDNSKP-----NDK----- 2227
 QY 2561 SKFYOEKRVCPVPRRHHMLRLDEIKERLDSNYLKMVRRTARNBGIDIIKFNSE 2620
 2228 -----RINSISAE-----DIEKRSNT 2245
 QY 2621 NGCAMPID--TMKYSFADIGIVRGTDMLRGVLPPEILKYVEFYIYKMKNNK 2678
 2246 QDVYMS-VSDSNTNGNKFYDLNDCIK-----GIFKGI----- 2277
 QY 2679 GRNKYNDVQTFRSAMMDANRDKIMKAMTCAKAPADAKLFRGRMDGFEITL-----IODK 2734
 2278 -----REDYWK-----GEYGVYDICTLEKTNNERY 2304
 QY 2735 GHRDPEVDYIPORF---RWMTWSEYCYKALMELEKFKKSCDCKTSDRCKNDYDEN 2791
 2305 SAKENDNKNQIILIRVLFKRWL-----ESFLEYNKINDKISHC-----MKND-- 2347
 QY 2792 KCEQCKTROEYKFNFLYKMKSLFDIOGNKRYEOPITYKISTYDHYOVNFOVKLTFKS 2851
 2348 KSPJCINGCONKCNVEKWKLEKKSSEWKVREY-----INOY----- 2385
 QY 2852 ECSVESFSEYLETSCINLKNFENDGSSNIRYAFEEETPKSYKEACSTLPKKNPLDC 2911
 2386 -----RDKNSN---EAPF--VKSLEFLITIPID-----V 2409
 QY 2912 PTDQNK-DGCKELOTFPFCSKNDYDNLDMMNAYLVANSDDNKGYLIPRRRRHLCTRPT 2970
 2410 VTDKGRHDSLTOLKRLKLCSEKSEN-----SNEKDYV-----LC-- 2445
 QY 2971 TAYNRRKQDEIILKKLLTSAFSOGOLGOKYSEELCEAKKYSADVSIDIIGKIDM 3030
 2446 -----LTK----- 2449
 QY 3031 DTSISEKIKIFETSNEATENRKTWENNRRQIMHAMLCGYKATSKVTLDEGMCOLPKD 3090
 2450 -----LEDKAKN-----C--KD 2459
 QY 3091 EETNOFLRWLEFYDYCYTROYKYLKDYOEKC--KSD--OLKCDTECNKCEGYVYK 3150
 2460 QASGE-----PCQUTSENDDDEDIIEE----- 2483
 QY 3151 LNLINKTMMNLKIKYKOLKDSGNDKNDKPSDENNOYSIKSDSOCALNELNINEIVG 3210

```

Db 2484 -----ENPVEAPNT-----CPKVEPEPVV--EBEKC-----DLAEPASK 2516
QY 3211 TKNNENNEFKVETLKLYGLYFVEDETHKNHVLDCGNIKKEEOYVRPKALFFPHVDSFY 3270
Db 2517 ESSTEENSGE-----GSMSEONPKRKEPEEPPEPPPSSETDP-----2553
QY 3271 QAPLESTHRAQVD-PKN---DIKSSIYVIVSALGLIALHFMKKKSSVDLRLRLNT 3326
Db 2554 -PPAPPTIOPSQADQPTNSISDILSTIPGIALALSIYFLFKKTKSSVDLRLVNT 2612
QY 3327 POGEGAPLTESKRRYIPYSSGPCKCTYTYMGGDTSCGDEDXKMWDLSSDITSSSEYE 3386
Db 2613 PKGEGYPTLKSSNRKPYPSADRKTKGTYTMEKDS--DSGHYED--TDTVTSSESEYE 2668
QY 3387 ELDINDIYVPGSPKRYTLIEVLEPSKRDIPSD--DTPS-----NDPEPTNRFTIDDEW 3437
Db 2669 ELDINDIYVPGSPKRYTLIEVLEPSKRDIPSD--DTPS-----NDPEPTNRFTIDDEW 2728
QY 3438 NELKHDSVSOYLPTNEN---NNYKADIPMTEPTLYSDNPEEKPFITSHDRDLTYG 3494
Db 2729 NTLKDEFISOYLOSQEPNDVNDYTSNGNSSTNTNITTTSRHNVYEEKPFIMSHDRMLYTG 2788
QY 3495 KEISYNNINNSTNTNDIPNARNDSYRGIDILINDSL 3530
Db 2789 EEINYNVNM-VNTMDIPINRDNVNYSGIDLINDAL 2823

RESULT 6
Q26030 PRELIMINARY: PRT: 3026 AA.
AC 026030:
DT 01-NOV-1996 (TREMBLREL. 01, Created)
DT 01-NOV-1998 (TREMBLREL. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLREL. 19, Last annotation update)
DE Variant surface protein (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE OF 2044-2922 FROM N.A.
RC STRAIN=IT 4/25/5;
RX MEDLINE=95330813; PubMed=7606788;
RA Peterson D.S., Ravetch J.A., Wellens T.E.,
RA "The large diverse gene family var encodes proteins involved in
RT cytoadherence and antigenic variation of Plasmodium falciparum-
RT infected erythrocytes."
Cell 82:89-100(1995).
[12]
RC SEQUENCE OF 2044-2922 FROM N.A.
RA STRAIN=IT 4/25/5;
RA Smith J.D., Chithis C.E., Craig A.G., Roberts D.J.,
RA Hudson-Taylor D.E., Peterson D.S., Pinches R., Newbold C.I.,
RA Miller L.H.;
RT "Switches in the expression of Plasmodium falciparum var genes
RT correlate with changes in antigenic and cytoadherent phenotypes of
RT infected erythrocytes."
Cell 0:0-0(1995).
[13]
RN SEQUENCE FROM N.A.
RP STRAIN=IT 4/25/5;
RX MEDLINE=99094502;
RA Smith J.D., Kyes S., Craig A.G., Fagan T., Hudson-Taylor D.,
RA "Analysis of adhesive domains from the A4VAR Plasmodium falciparum
RT erythrocyte membrane protein-1 identifies a CD36 binding domain."
Mol. Biochem. Parasitol. 97:133-148(1998).
[14]
RN SEQUENCE FROM N.A.
RP STRAIN=IT 4/25/5;
RA Smith J.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.

```

```

RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=IT 4/25/5;
RA Kyes S., Smith J.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: L42244; AAD03351.1;
DR InterPro: IPR004258; PFEMP.
DR Pfam: PF03011; PFEMP. 2.
FT NON_TER 3026
SQ SEQUENCE 3026 AA; 348124 MW; EC0A971A82193788 CRC64;

Query Match
Best Local Similarity 20.7%; Score 4010; DB 5; Length 3026;
Matches 1095; Conservative 480; Mismatches 1151; Indels 954; Gaps 127;

QY 30 ESHKSAENVLYRYAKNI-RHPSRYAKEHVDSLKGLTFAERGGSPSTVNNKNNYYPC 88
Db 13 EDDKDAKHVLDISGEIKYKVKVANYSSQKGLSLNAIFENEPKGGQTEND-----PC 67
QY 89 NLDHREHNLKRDVNLNHPCHGREQNPFDEDESECGN-KIRNYKKNDAIACAPRRR 147
Db 68 KLYEYHTNVTKGH-GREHPCKRKTGKRFSDVGGGECNRRKIKD--SKNNGACAPYRRL 124
QY 148 HMCDEKLEALNDINTQNIHDLGLNVLVTAKEYGESIYVNNHPRK-----GTSDACTAL 199
Db 125 HLCVARNLENISALDKINDTLLADVCLALHEGOSTIDYFKYQAQYASSFSQSICTML 184
QY 200 ARSPADIDIVRGIDMEKPNVHDK--VETGLREYFKIHGDMED-EYKNDYNPDGSGNY 256
Db 185 ARSPADIDIDIRGKDLTYGNKKEKIDLEKNLKFGLKYEKLTDPRAKDHKKEDPNNFF 244
QY 257 KLRZAMNWNKMYWEATITCDASYKSGFPMOSESENTPLFSNPKGHHQGR-----VPT 309
Db 245 QLRZAMNWNKMYWEATITCDASYKSGFPMOSESENTPLFSNPKGHHQGR-----VPT 304
QY 310 NLDYVPOYLRFDEMGEEFKRRNIKILKYVDSGR---NDKERLYCSHGHQDTTITW 365
Db 305 YEDYVPOYLRFDEMGEEFKRRNIKILKYVDSGR---NDKERLYCSHGHQDTTITW 363
QY 366 KGIHLNCKDCSTCKKVFPEWVLGNOEAFKKOKEYEKEIOSYSLNDKVV---NNT 421
Db 364 KHKIAMDACGCFSCSDPKRWIAKODEFEKKNNYTBKIK---KNDYTTITTEGTL 420
QY 422 NSEYKOFYEKLKTOYATNDTFLNLNEGKYCGK---GLPEKDTLTFNSDDGIEY 477
Db 421 NMYRKDPFYKHLER-KKTYDAFLNLNKEKCKNHPEVEBGKKYIDFDNIE---TFS 476
QY 478 RSEYCOYCPDGCYKCDGIKYTHKSDNDRERVN-----EDYKPPGVKPTNITYLSG 530
Db 477 HTEYREPCPWCGIE-----EOKDGKMRINDHSACKEEELTPKENAKYTKINVLTSG 529
QY 531 NEQDITOKLENCNSSTNYKDKNN-----OKMCCYKRD-----565
Db 530 EGHHDIAKRLKECTYKONGGSDGCGGSDSLCEPWCYQDPLERKVGGEVDKIK 589
QY 566 -----NINCKLEQNTENNDNPKIISFNFFELWYTLRLPTIKMNDLKTCIN-NTT 618
Db 590 GAGGICIFEKKGEKRYK-----KQTFNNFNFVAHVAKSIDMKRQULKLSLEDL 643
QY 619 THCIDECNRNCLCFDRVYKKEEENSIKKLPYKKKN-IQOYSYNNINLFGYEFKRVND 677
Db 644 KKECKGKSGNCEFFKWKIEKKEKWKVDOFNKQDPLFKHYLVLETLLEYVRENIO 703
QY 678 KLDKDEAKMELMENIKRKKNEFSNLENNRDYLENAITELLHLKFTATIC-----KDNK 732
Db 704 KAGGDLSIOEMKMKIKENKONKNTKD-----EDLADVLFHEKEADCDIHEDDD 759
QY 733 TNACETSHNATNCPVKGPGGTQPTKNIKEIAQYKRSAYEARNR--GLHKLGAHE 790
Db 760 DDECEVEIEKIPNNPC-----SGTRHRAVKNVAADYRAARQOLNRRBAGRTLAADSO 815
QY 791 GIYKRGRRKDFKDNLCRIMIKHSNNLTFNSGPDGCKTGIGIOTRFVYGVEMEVDPEH 850

```

Db 816 GHYNGKANSVLKD-VCDITNOYSNA-IGDSKDPCKGKDG-----FKICHTPTNTYVK 867
 Qy 851 MKRDEYIMPRRRRIHCSNLEHOTDHPNGNIVDDLNNNSFLGVLLSAKYEANKI 910
 Db 868 KTSYKDVFLPPRREHMCJSLNLENDVDYTNVN-----VANKFLYVLLLSANKAEKI 922
 Qy 911 IMKYK---KNLKGPEKVTDRKHOTTICRAIRYSPADIGDIIIRGDMERNNGDMYKLOG 967
 Db 923 KQRYEPNQNHNHKKR-----CRALKSPFADIGDIIIRGDMERNNGDMYKLOG 970
 Qy 968 HLEIYVGNHKSLSKGNKNDYNDAPKYLKRENNMWEARAVWEMKCDIYLNKSKSH 1027
 Db 971 NLVTTIFGKJAKVORKGIDISTYNTDGNHQLKEDMWEARNRROVYKAKCALK----- 1022
 Qy 1028 OSTOSSYCGSDHTPLDDYIPQRLRMTEMAMWYCVQKREYDKLEKCEKCKDNNGOG 1087
 Db 1023 ---GKINGC---ATPYDYDIPQRLRMTEMAMWYCVQKREYDKLEKCEKCKDNNGOG 1072
 Qy 1088 CKRESGTGTCTEACNEVNDIIGLKEBWNIIISDKY--KELHEQOMSVSNGIEASST 1145
 Db 1073 ICNKYKDCDAKCTEACKEKTKRIQPKKDMKLELEYALSYH----- 1115
 Qy 1146 AKNHIDR-----WIEFLSELYQO--NGKSNKSGTSDSAVIGNTTYENVAY 1193
 Db 1116 AKNDSRRMAFGGTDPDYQOYVHFPEKLOEALIKSSITSKRKRSTDAITPTPTPYSTAAG 1175
 Qy 1194 LHDYGNFDOQSONEFCDEKSDGKNEKYAFBDRPODHDAGCGSKSGPTRVOIKTKR 1253
 Db 1176 IHOEIGNACQJOIKHEDCNKD---KYFERKPRDHOEACGTENVE-----KPK 1222
 Qy 1254 ABEKDE---CTYNDILKENDKOVEDCHKKNNGTPD---WQC---GNINLEVPD 1303
 Db 1223 KEEDIKDLVCAIVKALTATDNLTKA--COOK---YGHPRHMGKCISETTKSSDSG 1277
 Qy 1304 RYCMPEPRORLKVHFLAN-----DNKIKLOSQVNLKEA 1337
 Db 1278 SLICVPRRRKLVVTLTKAELEATEBPTSPQAGGATLTPPATASQAARKGSL--LLRA 1335
 Qy 1338 FTKSAAEETFFSWYKXSKDGBENLDELKELKEGKIPPAFLRSFYFGDYRDLFGT--- 1394
 Db 1336 FIOSAVERFEFLMHKKY--MDNNGGDAEDLKEKEIPEEFKQMFYTLADYRDIQICVBD 1394
 Qy 1395 ---DISGHEGSKLKE---QIDSLFKNGDOKSPNG-----KTQDMWTEHSHL 1438
 Db 1395 VIKALEASSDNKSGNNIKKISDKIEAILKQSGSLGGLVTPPNVKNKRTWMDONAKHI 1454
 Qy 1439 WEMALCALV---KIGAK--KDDFTEN-----YGNVNVKPS----- 1468
 Db 1455 WHGMVICALYKEDTGAKKGTSTIDPTAYGKIMDMGCKKPKEDKYDKIYKISSVPSN 1514
 Qy 1469 --DKSTLEEFAPKPOFLRMLEWTDYCYTQOKYLKVOEKC-----SNDQL 1515
 Db 1515 VPSGDTKEEFSSRPFFEMLEWGEFPRKRKDKLEKIEKECYGLNYRQIHYCSGDGH 1574
 Qy 1516 KCD-----TECNKCKEDYVKY--KKKKEMIPQDKYKDERDKRPRRPHI 1559
 Db 1575 ICCKTDSNNMTFIDHCPRCLEKCIKRYKWLKKEKEPHNOKNNKEKFN--DLKER 1631
 Qy 1560 GWAVTDTGTNATDYLNRKFTASCGD--KPGSASVQOR--IOLLEKOAYDADKHC--GCTR 1616
 Db 1632 G-----YSSFN-----NFLASLNCHKGEHNRDKNKIEFNHNTKTPSPSYCACPEY 1679
 Qy 1617 FLENDKYTNISSKCKQGLVEANTGAIRKMONKGNPNYNNLKELEEDVLFPSRLRLICE 1676
 Db 1680 YGVKCNK-----KNGCEETIHKTDLNG-----QNDNNYTDIKVLYD----- 1716
 Qy 1677 HALDGYTDPYKDEGLKRLMEVAATGYNLGOYKKEKKEE-----KITSDAKH 1729
 Db 1717 ---RKGSSNDEELKNNVNNNTSLFKDSSV-----QYWKQCKNNEVDQCIITNFDLIDIK 1767
 Qy 1730 YSEYVPCASMKYSFYDLIDIIIGIDNLEDEKQTEENLKPIPNKNGTSVSGKSDSTTG 1789
 Db 1768 Y-----MEFNVPFQRLRYFVHDYINILKDKIKPCK-----TDEKSNKCIINGC 1811

Qy 1790 POSTAKRFFPMNENKCECVNNAAMICGYKRGDDGNSG-----NSARSDBLKCKGCV 1839
 Db 1812 KGLCEVKKMLKQKODEMNKINDHYEKNKSLGYGIPHWKSYFVQLVEDDYKKAQOV 1871
 Qy 1840 PSDDDYPMKKNDEGTAYOFLRMFAWGEDFCKHKKELKLVGACNDYTCGDNEDEKRRK 1899
 Db 1872 IED-----ENBRKKTWCGTDCGVECTNEETKREK 1899
 Qy 1900 CTDACTQYKKTISEMKPOYEKOIKKYGKNDKIYSEHPVAKADEADREYLDQKIKCN 1959
 Db 1900 ---DETNLKIKLOEKIESCOTQHPNG--KYRCDIPIPHSD-----ETLDEDTDTTDD 1950
 Qy 1960 KSGDCBYK-----CMKDVSOTRLDQNSQMPASLDEDEKEVEGKCNQVPRKPPVRRE 2014
 Db 1951 DMSDKIYDRKPPPCPDVEDTET-----EKXK-----VLGCP----- 1984
 Qy 2015 TSPRVSLSKATASKKEAKTAPPTQPKVENLTTEMAQOTRRAAQTTRKSTATT 2074
 Db 1985 ----- 1984
 Qy 2075 TESDVGTAKILSNKPDRCGIEGCPKTY---GOYPMKCIYVSKRENEGICMPRR 2131
 Db 1985 ---DACEIYGEIL--NGQDGTKEIEBCNTKYTPKNDYPCGMC--TDKVINREBSCMPRR 2039
 Qy 2132 KILCINNIOYLVNTEKRDNDIKAEAFIKCAALETOFLMKY-----IEN 2177
 Db 2040 OKLCIHNLBHL---SEKATELEKRAFIKCAALETFLMDKYKEDKDEKTEGGISD 2096
 Qy 2178 P-AAENELONGTIPDEFKRIIMTYTDYKDMFPGTISDK--KIITVTSVTTILNENK 2235
 Db 2097 PDDPKLLEGITPEDEKQMEFYTDYRDFLEGTISGHEGSAKGIKIDSLFRNGQO 2156
 Qy 2236 KRODKKDELRKIFWEKKNKFIWEGMIGLYGLVLDENEKERIRNY--QYNDMTKLTPS 2294
 Db 2157 KPSGKTPPE---WANDGPDIMKMGVGLSHHINGN--KDLRKNLNDNNKTYTISK 2211
 Qy 2295 LEEFYKPOFLMFTWAEBCNKRREOLIKLEAGKEYECNGSNDGKTOEACAVTYO 2354
 Db 2212 LEDFASRPQFLRMFIEMGQFCREBYVKINOLKTGCNEVEGSGQENKKEACKNACEAVK 2271
 Qy 2355 NFKKKTETEYERQREKFKDKGKKKVDYPTROLEKATCAHEYLNMKLKELCGKDS 2414
 Db 2272 SWLKMDKDYEDQTAFFDKDKDKF--DETSAVDVAASVSEYILOELKINLCITGDDA 2330
 Qy 2415 CMQKPSQPLPKTQOSSDANDMPESLDYVPEEF--NKCEPELSKSGSNHTKKTTEK 2473
 Db 2331 CMKPSAQ---DETTELLOGNFPRAMDYPREIGERCK-----AIPS 2371
 Qy 2474 IPNVCYKAAVYLSKAEENNDITLKEKFIPIESTEKESKSNWTNNPCDPKPYAPDK 2533
 Db 2372 EPMSCVEQJAKHLREKAEKNVKEYESS---LKGTPAK--SKNDCTRID--EALKGNGSK 2424
 Qy 2534 YIGR-----RNPCKRRE---NREKDYEMKCYNSFRYOQKKNVCPPREHHCN 2583
 Db 2425 IINKSLDSTPASCQOSEADATDRUKIGOWE---NKNINGETLYVPPRRKDCEND 2481
 Qy 2584 LDEIKIERLKDSDNYLKVARTARNGIDIIKFNSENGCANPICDTJKYSPADGDIY 2643
 Db 2482 LKNIQENVEODSLSLEKIQHAKNKGIDILKLNQODQONAPSEIDAKYSPADGDIY 2541
 Qy 2644 RGTMDLRLIGYLPPEBIKLYVEFYLYGKWRNKNRGRNRYNDVOTFRSAMWANDRDIK 2703
 Db 2542 RGRSKID-----PTNNKITEKELQIFKOIODNASTSKM--ELPELREKWMANDRREYVN 2595
 Qy 2704 AMTCAPADEAKLFRKGRMGFERITLIOD-----KCGHDDPPVDYIYIQRPMWTESE 2758
 Db 2596 AMTCVAPNDHAHLKKNKNNPNNSQIIASOTEOYTKCSHSEPPDYIYIDERRHFLDEWSE 2655
 Qy 2759 YYCKALMELEKFKKSGDHC--KTSDRCKNDYDENKCEQKTRCOEYKNFYLKWSLFDIO 2817
 Db 2656 YYCKALKEKNDEKKNCSKICSGAKCEKKEDEKCEKCNCKEYKKNYVDKQSOSEFDO 2715

QY 2818 SAKYELREPIYTKISTYDHON--FVOKLTKFKSECS-VESFSEYHETSCLMYKEN 2874
Db 2716 NOYKYLITODRTHGSPTRARNPSIEFTOKL---EDSCNDPYGADKYLDISTHCTDYKTS 2772
QY 2875 ENDGSSNIRIYAFETPKSKYKEACSCOTLPBKNPLDNCPTPDOKNGCKELOTFTPCSKNDY 2934
Db 2773 ETD--SRESYAFSPYPKDYKCKCKV-----NTPt-----SND- 2806
QY 2935 DNNLDMNMYLVNSSDNDKGVLIIPRRRHLCPTITAYNRYKGDKEILKKLLTSAFSQ 2994
Db 2807 -----PK-----S 2809
QY 2995 GOLLGOKYKSEELCEAKKYIADYSDIITKGTMDADTSLSEKIKIETFSNEATEENRT 3054
Db 2810 PSLIGSPF----- 2817
QY 3055 WNNRRQIWHAMLCGYKIATSKVTLDEGMCOLPKDETNOFLRWLEMAKQACKKKHV 3114
Db 2818 -----FLPK----- 2821
QY 3115 SDSLTKCPRSNEDNEASELLRQPCQNDIRKYISLNLINKNTMENLNIKKYKOLKQSS 3174
Db 2822 ---KPK-----MKFYPRIGIGVLAHPFINM----- 2842
QY 3175 GNIDNKPSEENVSYIKSKDSOCALENDINELIYGTKNENNEKEVILKLYPGLIYVE 3234
Db 2843 -----VADPTIHEFTVAKTFENNAVQF----- 2864
QY 3235 DETHKNHVLGNIKKEEQTVPKALYFTPHVDSFYQAPLFTSHRAVQYPKNDILKSI 3294
Db 2865 -----HINPKTDVAP-----TKNt-----LNEVLPSNI 2889
QY 3295 SVYIVSAIGLILHMKKKKSSVDLRLNIPQEGYGMPTLESNNRYIPYSGPYKQK- 3353
Db 2890 PYGIMALASGLAFLELKKTKYHVDLFVYINIPKSDYDIPKLSNRYIPYSGPYKRGNG 2949
QY 3354 TYIYNE--GDTSGDEDKYMDLSSDIT--SSESEYEELDINDIYVGPSPKYTLEIYVE 3410
Db 2950 TPTLKEIYGTDSGYDHY-----SDITSSSEYEELDINDIYHVLNLTITLLEVLE 3003
QY 3411 PSKR---DIPSD--DTPSN 3424
Db 3004 PSKLSGNTIIPSGKNTPSD 3023

RESULT 7
Q26031 PRELIMINARY; PRT; 3078 AA.
Q26031: 01-NOV-1996 (TREMblrel. 01, Created)
01-NOV-1996 (TREMblrel. 01, Last sequence update)
01-DEC-2001 (TREMblrel. 19, Last annotation update)
Db Variant-specific surface protein.
GN VAR-1.
OS Plasmidium falciparum (isolate Dd2).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=57267;
RN [1]
RP SEQUENCE FROM N.A.
RC STPAIRN-DD2;
EX MEDLINE=95330813; PubMed=7605788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
RA Peterson D.S., Ravetch J.A., Wellens T.E.;
RT "The large diverse gene family var encodes proteins involved in
RT cytoadherence and antigenic variation of Plasmidium falciparum-
RT infected erythrocytes.";
RL Cell 82:89-100(1995).
DR EMBL; LA0608; AAA75396.1;
DR InterPro; IPR004258; PFEMP.
DR Pfam; PF03011; PFEMP.2;
SQ SEQUENCE 3078 AA; 349297 MW; C8037C2B03CCD7C3 CRC64;

Query Match 18.0%; Score 3489; DB 5; Length 3078;

Best Local Similarity 26.7%; Pred. No. 2.7e-160;
Matches 1054; Conservative 472; Mismatches 934; Indels 1490; Gaps 148;
QY 9 IIKGNMASSLEGGAKSPLIKESHKSARNVLYERAKNIRHPSK-YAKHVDLKGDLTKA 67
Db 4 LAKMGPKEAAGDDIED-----ESAKHMFDRICKDYDYKVEAKRGGLGRSLSEA 56
QY 68 EFRGSPPTPVNKHNYYYPCNLDHKEHTNLRDYDYNLRHPCHGHEONRDEDESEC-G 126
Db 57 KEKNESEDQPED-----PCDLDRHYFTNV---TTNINPCADRSDFRFSDEGCGCTH 108
QY 127 NKIRYKRNDALACAPRRRHCKNLEALNDINTONIHLLGLVLTAYEESSTVNN 186
Db 109 NRIKDSQGDMDKACAPRRLHVCQNLQEOLEPIKITHTHLLVDYCAAAFFEGOSTIOD 168
QY 187 HP-HKGT-----SDACTALASFPADIGDIVGIDMFKNVND-----KVETGLREVERKI 235
Db 169 YPKYQATYGDSPSOICTYMLARSFADIGDIVGRDLVGLNPOEIKOROOLENNLTIFGI 228
QY 236 HD---GMEDEVKNYNPDSGNYKRLREAMNVRNRYWEALITGDAKSKGYFQSESNT 292
Db 229 YEKINGAARFGND--PE---FFRLREDWNTANREYVWKAITGNA-WGNTYF-HATCNR 280
QY 293 PLFSNPKCGHKGKVPNTLDYVQYLRFWDEMGEEFCRKNRIKLYKVDSCR---NDKE 348
Db 281 GERTEGYCRNDQDVPYFYFDYVQYLRFWEEMADFCRKNRIKLYKVDSCR---NDKE 340
QY 349 RLYCSHNGHCTTYIMKGIHLNCKTDCSTCKKVEEVLGNOQEAFFKQEKERYEIQ 408
Db 341 R-YCSRNQYDEKTKRALIGRLRYGKOCISCLYACNPVYDWINNOKEQDFDKKYYDEIK 399
QY 409 SY-----LSNDNKFVNINSE-YKQPEYKLEKTOYATNPTFLNLBEGYK- 454
Db 400 KYENGASGSGNQKRDAGTTTNTYDGEKRYDLANKSEKRTYDLEKISNEEICLYK 459
QY 455 --KGLGEKDIYTTNSA-----DKGIFRYSEYQVCPDGVKCDGIKYTH-KSND 504
Db 460 DEEGGIDTFKNVNSDSTSGAGTNEBQGTFFYRSKYQCPQPGYK---KVNNGSSNE 515
QY 505 RERYNN-----EDYKPPWGVKPTNIVLYSGNQGGITQKLEFCN-----SS 547
Db 516 WEKNNNGKCKSGKLYEPKPKREGITITILSGKHDDIEKLNKFCDEKNGDTINSGSG 575
QY 548 TNYKDKNN-----QKWEQYKDNINRCKLRONTIENDNPK-----II----- 586
Db 576 TGGSGGNSGRQELYEEMKC-YKGEDVYKVGHDDEDEYENKMGGLCILANKQKKE 634
QY 587 -----SFHNFELMYTLLDITKANDKLTCTI-NNTTHG-IDECNRNC 629
Db 635 EGGNTSEKPEDELOKTYNPFYFVWVAMHLKDSIHWKKILORCLONGNRILKGNKCNDC 694
QY 630 LCEDRWYKQKEEEMNSIKKLFTKKNNIQ-----QSYSNINLEGEY- 672
Db 695 ECFRWITQKKDEMGRVQH-FTONIKRGSGSDNTAELPFRHDVLYOYNLOEELKGD 753
QY 673 -----FKVVDKLDKDEAKKELMENIKRKNFSLNNRD-----YLENATELL 718
Db 754 SEDASEEKSNSIDAEEA-----EELKHLREITIEEDNNOEASVGGVTEQGNIMDKL 807
QY 719 DNLKETATIC-----KDNNTNEACETSHNNTTNCVPRGTO-PTYNIKEIQ 766
Db 808 NYEKDEADLCLEHDEEKEKGGDNECEENRYNRCSESENKRRPYLVANKVAQ 867
QY 767 YFKRSAYEARNGLKLKGAHEGILYKRGRRKDEKMLCRIMKHSNRNLQFSNGPCD 826
Db 868 MHHKAKTQLASRAGRSALRGDISLAQFKNGRNSITLKGQICKINENYSDSGNSGSPCT 927
QY 827 GKGTGCGIQTRFYVGEWEYDEPHMKKDEHDVIMPRRHICTSNLEHLQTDHPHNGNI 886
Db 928 GK-DGDHGVYRMHIGTEMENIEBCKQTSYKKNVLPFRBEMCTSNLENDY-----GSV 980
QY 887 V-DDLVNNSFLGDVLLSAKYEANKIIRMKKRNKLGPEYVDP---KQITTCAIRS 942

QY	2898	CGCTGATPSKPNFIDNCPDQNKQCEKLEDTFPCSKANDYDNNLDMWNVLYVLSNSDDNKGL	2957
Db	2427	CCGSASA-----NENKNG-----	2440CN
QY	2958	IPPRRHLCPTPIYAYNRKDGKEILKKLLTSFASOGLLGQKYSSEELCFEAMKYSY	3017
Db	2441	-----	2440CN
QY	3018	ADYSDLIKGTDMQDTSLEKIKKIFETSNATENRKTWNENNRQITWMLMGYKIAISK	307773
Db	2441	-EYKAI--DOMLKKLKD*-	2457
QY	3078	VTLDEGWCQLPKEDEETNOFLMILEMAKOACKERKHVSLSLTKC-----PRSNEDNFEA	313232
Db	2458	-----GEC-----EKHHQHS-DTEGCDTPQPTLED-----	246833
QY	3133	SELLAQPCQNDIRKYSILNLINTEMNINITYKOLQOOSGNDINKPSEENVOSTYKS	319292
Db	2484	-ETL-----DDIETEERKKMMKICEN-LKTAQOEDE-GGCYPAENSEEPA-----A	253030
Db	3193	KDSOCALANDINEIVTGKNNENNEPEYVKLYGGLYFVEDETHKHNVDGNKKEEQ	3252
Db	2531	TDS-----GKEPEPTPLK-----PEDEA	2550
QY	3253	TYRKALYFTFPHVDSTYQAPLSTHRAVDKPN-----DILSSISVYIVSALG	330333
Db	2551	VEEPP-----PPPOEKAP-----IPQCPPTPEPTOLLNPNHYLALVYSLTAMSVG	259999
QY	3304	L-----IALHKKKKSSV-DLRLINIPQEGMPTLESKNRYPYSGPKKGTYYM	3358
Db	2600	IGFAFTEFYFLKKTSSVGNLQIILQIPRSDVIDTKLSPKRYIPYISGKRGKRYLYL	2659
QY	3359	EGDT-----SGDEDKYMWLSSDITSESEYEELINDIYVPSPKYKTLIEVLEPSKR	3415
Db	2660	EGDSTDSGYTDH-----SDTSSSEYEEMDINDIYVPSPKYKTLIEVLEPSGNN	2713
QY	3416	IPSD-----DPFS-----NDTPRNFIDENNELKHDFVSOYL	3449
Db	2714	TTASGNNTTASGNTTASGNTSPDQNOIQNDGIPSKYITDENNOULDEFTISQYLOSE	2773
QY	3450	PNTEPNNNKYSADIPNTEPNTLYSDNPEKPFITSIHDRDLYTKETSY-----	3499
Db	2774	PNTEP--NMLGYVNDNTHPTTSH-HAVEKEKPFIMSIHORNLFSGEENYVDMFNSGNPI	2830
QY	3500	NINNSTN-----TNNDIPNANSDSYRGDILINDISLVANILTY--KMK	3541
Db	2831	NISDTSMSDLSNNSHPINDKNDLISGIDILNDLISGHNHIDYDEMLK	2880

Query Match	17.2%;	Score 3328.5;	DB 5;	Length 2527;
Best Local Similarity	27.6%;	Pred. No. 1.2e-152;		
Matches 982;	Conservative 421;	Mismatches 913;	Indels 1241;	Gaps 134;

QY 31 SHRSANVL-----EYAAINIRHPKAYAEHNDISLKGOLTAEPFGGP---STVPKNHNY 82
I : I :
Db 13 SAADAELMDMGEVYKKREKVNA---ANDPEFKLGTLSQAETFEAKEDOOTGN---- 65
QY 83 YYPFONDLKHEHTNLRKDYNVLRPHCHGREQONRDEDEESBCG-NKTRNYKRKNDAIAC 141
:
Db 66 ---PCLFYOWHTAATR---GKNYPCBFTGETKEPRSEVSGBGCDBECKIKDNKGKEG-AC 111
QY 142 APPRRHMOCKMLNLINDINPONIHDDLGNLYTKAYGESIVANH-PHKTSQA---CT 197
I : I :
Db 117 APYRLHLICVRLENISALDKITNDPTLDADLCALHEGAALISADHGQOQTNSOLCT 176
QY 198 ALARSPADIGDITVRGIDMK-PNVHDKYETGI.REVEFKRIHOGME-----DEVKNDNP 250
I : I : I :
Db 177 MLARSPADIGDIIIRKGDILYRGNGNKDKLEENLKIFGNITYKDVYKGGKANVALKTRE-D 233
QY 251 GSGNTYKLREAAWNNRNKVMWEATLCDASYSXS---YMOSSESTPLFSNPKCGHKQGV 307
I : I : I : I : I :
Db 236 AIGNTYKLREDMMANROEIFMALTCDA--PDGVNHFRKCSMGOSHVDKCYCLNGDP 293
QY 308 PPNLDIYPOLYRWEDWEGEBECRKRNIAKLKYKVDSCRNDK-ERLYCSHNGHCYTYYMK 366
I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 294 PIYFYFVYPOLYRWFEEMAEDEFCETKRHKLOVAIKICRDSGNDRYCDLNGDYDTKTARG 355
QY 367 GJLHLDNKCSTDSTCKAFYEVLWGNOZAPFKOKREYKEL-----OSTLSDMNFPVN 419
I : I :
Db 354 NKRFSDNECYKSLPCHDFHVIMIDNQCFEFKOKKTTAEIKABEEKKESIITIGDTIN 413
QY 420 NINSEYQOFYKLEKETOTAIANDTFLNLINLGKYCK---GLRGEKDIYFTNSADDKGI 475
I : I :
Db 414 NI---YVOEFYQOL-QSSYKDVEDFLOKLINEERJOCKRPYGTETAPDVFYKCTKEK-T 468
QY 476 FYRSEYCVCPGCCGYCOCIGIKYUTHSDDNRERV-NEDKRPAGVYPIJTI---YLXG 530
I : I : I : I :
Db 469 FSRTEYCRACPMGCIK-----KKAOSTWRILDMNKCYKEKTTYKKAITPIDVLPD 521
QY 531 NBOGIIOTKLEPFCCSSINTYODKNNQ-KWCZYK-----DENINRCLE---QNT 576
I : I :
Db 522 KSOSRILDKYVNFCD-DINGNSNGOKCEKWICYKKNNEKBDVANGDSNI-CVLQDNKOPT 578
QY 577 EINNNDPKIISFHNEFELMWYTLRDITKNMDDKJCTGINNTTHCID-ECONRJCLOERM 635
I : I :
Db 579 BEENDR-----SYNSPYKWVTYEM.LIDSIDRKTELKCTNNKRGCKAKKCKANNCKPERW 634
QY 636 VKQKEEENSITKLFYTKKNNIQOSYASNINNFLFBGYFPKYWDKIDKDEAKMTL-MENT 693
I : I :
Db 635 IEKKNEKEKLVDFEKPOPGF-DILSNNDLALKTLIDVEDILTINEJDEYKEVGSSEHI 692
QY 694 KRKKNEFNENLNNDYLE-----NATEILLDH.KETATICKONN-TNEAC---ETS 740
I :
Db 693 RKM-----LENEKNOBEAAVAVYKRNITIDLIEHLEDPADLCLOTHLEDCEPSDEDD 746
QY 741 HNA---TTNPCVVRGSGTOPTKNIKETAOTYFKRSAYEEARNRGLHR-LGRKHGTYKRG 796
I : I : I : I : I :
Db 747 HEEEVYVNPAPKPSGSGYPSLAN-EAHOHMRAAKTOLSGRGRRTLIKADASKGYKRG 804
QY 797 GRKDFDONLCIRMIKSNNRLGFSPNCPGCKGGOIORFVGVRGEWEDRPHMKHKE 856
I :
Db 805 GQGUTLNGIDICKIKNTTN-DISNGACACKG---DOSNERFIETGEWSVEDKTTSTS 860
QY 857 DYVMPRRRHCTSULEHOTDHPNLGNITVDLVVNSSFGDVLISAKYEANKIIRMYKE 916
I : I : I : I : I : I : I : : : : : : : : : : : : : : : : : :
Db 861 DVIYLPBRREHCTSULENDLVGSTENGKAI-----HSILGDVILITAKDMAEIIRKXS 915
QY 917 KNNLKGREVYLDPHQOTICORAIRYSPADIGDIIRGRDLERNRGOMVYKLOGHLENYFGNI 976
I :
Db 916 ONSIGD-----IDKXHOSICORAIRYSPADIGDIIRGRDLWNETGMNHLKGHEKFFKI 972

QY 977 HNSLKGNDKYNDNA---PKYLKRENMWEANRAKWEAMKGDIKYKDKSGHSTQSS 1033
Db 973 KEKLPET-ODKTYGDERKNNPPLYLLEBDWEARHOFWRMRCATP--KDK-----NM 1022
QY 1034 YGYSOHTPELDYIIPOLKMTWEMAEWYCKVOKREYDKLEKCECKDKONGOCSTRESG 1093
Db 1023 KC---NGIPLEDIIPOLKMTWEMAEWYCKVOKREYDKLEKCECKDKONGOCSTRESG 1079
QY 1094 YGCTCTEACNEYNIDILGMEQWNIISDKYELHBOAOMSVANSGLIENASTARKNHD- 1151
Db 1080 V--CKPAKOCKEYKIKMEEQWRIKSDKYNLLYLQAEFTARNAG---ATSPDGDIP 1134
QY 1152 -RNVIEFLSL---YOONGSKNSKSGTSDESANIGTNTYENGAUHLDTGNDDCOSQ 1207
Db 1135 YKQVVPFEELOKTISASAKRPKSTDTGTTPTLSPYTAAGYIHOELPMWGCISQT 1194
QY 1208 EFCDEKSDGDKNEK-YAFNRKPODHOGACGCKSGKPTROJTKKKAEKDECKTVND 1266
Db 1195 HFC-----GDNKKEAFNRKPDHDKVDCSEGER---LOYA-KKEGYLVEPCNYVKT 1245
QY 1267 LKENDGKQVEDCHPK-KNSN-GYPMOCGNINL-VEBPRVCMPPRRORLCHVFLANDN 1323
Db 1246 LKDKKNNSDICEIKRYKNGKNEKYPWMDCTRNKIKIGEAGAYMPPRRORLCHVFL 1301
QY 1324 EIKKLOSQVN--LKEAFISAAEFTFSWYIYKSKDEGN-----ELDKELKEK 1371
Db 1302 --KOLDOYDEKLRDAFISAAEFTLSWIKYEDKREKPEEASLDVEAOTOLDNGI 1359
QY 1372 IPAPAFRMEYTFEGDYRDLFGTIDISKGHSGSKLEKQIDSLFKNGDOKSPNGT---RQ 1428
Db 1360 IPEEPRKMEYTFEGDYRDLFGTIDISKGHSGSKLEKQIDSLFKNGDOKSPNGT---RQ 1416
QY 1429 EMTTSHETWEAMCALYKIGAKKODTE--NYGINNVKFS-DKSTTLEPAKRPQFLR 1485
Db 1417 VFMGTGKIDMEGLALEKITEKQMTFTDKHEYTFDKVFSGDKSBILEKFAOTQOFLR 1476
QY 1486 WLEWYDQCYTROKYLKDOVEKCKSNDOLKQTECNKCKCEDVYKMKKKEMIPQDKY 1345
Db 1477 WMTGMEHHCERKEKEVEILR-----TEC----- 1500
QY 1546 KDERDKRPDRHIGVWVTDTGTNATDYLNRRFTASCGDKPGSASVQHNIOLEKQAY 1605
Db 1501 ----- 1500
QY 1606 YDADKCGCTKELENDKYTNISSKDKGLVEANTGAIKQNKGNPNYNNLKELETV 1665
Db 1501 --AGC----- 1503
QY 1666 LFPRLRLICFHALDGNITDPEYKDENGELKRLMEVAATGYNLGOYKKEKKEKIKTS 1725
Db 1504 -----EVS----- 1507
QY 1726 DAHKYEVPPCSAMKYSFYDLNDIILGIDNLEDEKQTEENLKIFNNKGTSGKSGDS 1785
Db 1508 ----- 1507
QY 1786 TTGNPGSTARKFTWENKECVMNAMICGYKRGDDGNSGANSDEDLKCGSVPSDDY 1845
Db 1508 ----- 1507
QY 1846 PMGNRDEGTAQOFLMFAWEGEDFCKHKEKELEKLVGACNDYTCGDNEDKRRKCTDACT 1905
Db 1508 -----NDGSCCHKNGECCCKRECK 1527
QY 1906 QYKFISEMPPYOYKIKYGENKDIYSEHPAKDAEDAREYLDOLKTI---CENKSG 1962
Db 1528 KYKMLEKRENYDKQVKFKTDEG-YNDPDTIOSTEAYELGKLLKNTITFTSGTNG 1586
QY 1963 DCEYKCMKDVSTORLDGN--SOMPPASLDEPEVEBCKGCQYPRGPPRRKRETPSRVS 2021
Db 1587 DC--KCMQEKSKQSHDGSITDMMPESLDEPEVEBCKGCQYPRGPPRRKRETPSRVS 1639
QY 2022 LISKATASKREAKTADPTKQPKVENVLTTEMRQOTRRBRAAOOTRKRTSTATTTESDVT 2081

Db 1640 ----- 1647
QY 2082 MYKAILSNKPDNRSGLEGCPNPTYGQ---YPMKGTIVGSK--ENENGICMPPRRKLK 2135
Db 1648 IYVAIID--PNSGTGYDACEEKYGNKTAVPBGKCTTNKIEGKEDDVYCIPRRKRL 1705
QY 2136 INNIQYLVNTEENKRDNDIEAFIKCAITPOFLMLKYLE----- 2176
Db 1706 IKQKL-----EFSCKNDKELKRAFECAAVETFPAMHKKEDKKNKEEKESTYLVRLGL 1761
QY 2177 -----NPAENELONGTIPDEFKRIIMTYTGDKYDMFGTIDISNDKILITVNSVTI 2229
Db 1762 GKQOSSPEEBAORDLESGLIPDEFKRIIMTYTGDKYDMFGTIDISNDKILITVNSVTI 1819
QY 2230 LNENNKQKODKKDEELRKIFEMKNNKFEIWEKNGYIGTLYHLTDBENKEKIRDNV---OY 2285
Db 1820 F-----PNSVYKPDGQRETRWNNNAKDYNMICALST---DTKRIKIEGITYAOLTON 1870
QY 2286 NDMTKLT---PSLEFVRPOFLRWETEMAEFECNKRREKQDLKLEAGCK---EYEC 2335
Db 1871 NDYKNTYFDNGNTTLEIYATVPQIRSFEMAEFECNKRREKQDLKLEAGCK---EYEC 1930
QY 2336 NGSNDGKTQF-----CAEACVYQNFJAKKTEYERKREKFKDK----- 2375
Db 1931 SGGHDCTNEELKHNDFADSYCPDCKKACRKYNKWEKVEEYFNOKSKYKREHKTSS 1990
QY 2376 --DGKKYKDYPTSERDIEKATCAHEYLMLKLEKCGKDCSQMOQPSQLPRTQOSQS 2433
Db 1991 NDNKYKKEFYATSEG--KKSVDSPFLD-KLKE----- 2020
QY 2434 DANDPESLDYVPEEFNKCEPPELSKSGSMIHKKITEPKIPANCVEKAAVYLSKEAEN 2493
Db 2021 -----RSHC----- 2025
QY 2494 MDITKEKFIPISTKEKESKSWTNNNPQDPKPAADKYIGRNP-----GENRENR 2548
Db 2026 MD-TLEBKT-----DKN-----PLKTSSTYTC-KTCPLVYVOCRNSTDH- 2064
QY 2549 FKDYEMKCKYKNSKRYOKEKRVCPVPRRHEMCLRNIDEITERLKDSNYLLKAVRAR 2608
Db 2065 -----CIQNSG--NEK-----WEHA--LDITIKKAPRISINVOIMDR- 2099
QY 2609 EGIDIIKENSEGCMANPICDTMKTYSFADLGDIVRGDTMLRGIGYLPVEIKLYVEY 2668
Db 2100 -----RG-----OY 2103
QY 2669 IYKWRKNNKGRKYNDOYTFRSAMWDANKDIY-----KAMTCAPADAKLFKGRM 2721
Db 2104 I-----QEHSENSFKESRLKSV---REOKWECSEYVKKKDYCV---LKNF-KENI 2147
QY 2722 DGEFRLT---LIDDKCGHKDDPVDYIIPQFRMTEWSE--YCKALMELEKFKKSCDH 2777
Db 2148 DTDETITFKVLLN-----WLOPFIGYIIS-----KRIIDI 2179
QY 2778 CKTSDRCKNDYDENKCEQCTRCQOETKYNVLYKWSLFDIOSNNKYLEYOPITYKISTYD 2837
Db 2180 CTKKE-----EHTALEGCKSKCE---CIGK-----LQKTEWMD 2211
QY 2838 HVOFNQKLTFFSECVSEFSELYHETSKLWAKFPENDGSSNIRYFAEERPKSKYA 2897
Db 2212 EIKTHENKO-----NRDG-----YETAHVRNTE- 2237
QY 2898 CSCSLPSKNPLJDNCPDQNDGCKELOTFTFGSKNDYDNIMDNNAVLYLNSDNDNGYL 2957
Db 2238 -----KNMV-----OLKKV-----IDDLKLV- 2253
QY 2958 IPRRRRLCTRPITAYYKRGDKELKLLKLLTSAFSOGULGQYKSEELCEPAMKYSY 3017
Db 2254 ---KRIDDRYIYVLDACANMKE-TNKDVSJILSO---LKREIKPFENQPHETSPNY 2305
QY 3018 ADVSDIKGTDMDDTSLSEKIKFIERTSNATEENRKRKWMENNRRQIWHAMLCGYKIATSK 3077

```
Query Match      16.1%   Score 3130.   DB 5.3:   Length 2924;
Best Local Similarity 26.7% : Pred. No. 6.1e-13;
Matches 1005; Conservative 445; Mismatches 968; Indels 1350; Gaps 151.

QY      33 KSAKNLIERAKTI-----RHPSRYAKEHVDLSKGDITKAERFGSPSTVNNKHNYTYP 86
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
```

Db	13	KDAKHAIDRIGEEVYKKEVNAEKYKK-----ALKGNIDQEAIGTELASS-----	PN	60
Qy	87	PCMLDHEHTN-LRYDVNLRRPCHGREDNRDEDEESECG-NKIRNYKRRNDAL-ACAP	14	14
Db	61	PCFLVEDYIYNNRK-----RRRYPCANQRQYVRSDEGGCTFENRKIDSENNDSIGACAP	11	11
Qy	144	PRRHHODKMLLEALNDINPONTIDLLGNVLYAKYEGESIVNHN-PHSGT-----SDACT	19	19
Db	117	YRRLHCYDNLLEKMGTSYTK-HDLLLDVCAAKIEGSIKTHYKHELYTPDRKSQJCT	175	175
Qy	198	ALARSADIGDIYRGIDM-----KPNVHDVETGLREVERKIKHDGMEVDKNDYDPDG	251	251
Db	176	ILARSFADIGDIYRGKIDLYGDDEKEKERKLENNLLEIFEKIHEHNLGTQADADHYKKD	235	235
Qy	252	SGNYVILKREAMVNVNKNWEALITDADASYKSGYQMSNTPRLSPN-KGCHKQK----	306	306
Db	236	EENYVYOLRDEWWTANRSTYWKALITTHAGESBDRYFKTKCSEBMTDDCKRCDEGKNETN	295	295
Qy	307	-VPTNIDYVPOYLRFMDWEMGEBCRRNKIKLYKVDSCRNK--ERYLCSHNGDCTTTI	363	363
Db	296	EVPYVEDYVPOYLRFMEEAEDFCRRKKIKLEIATKNCRGEGNER-YCOLNGNCEETA	354	354
Qy	364	WKGKGIHLNCKTDDSTCKQVYEWLGNQDAFAKQKRYKELO-----SYLSDNKFVN	419	419
Db	355	RGAEIFYKQDDCHCKSVACDREYKVIDNOREFQRKKYKDEELINTKHTGTTIATGNKIN	414	414
Qy	420	NINSEYKQFYKLEKETOYATNDTFLNLLNGKYCK--GGLPGEKIDTTFNSADDKIEY	477	477
Db	415	NL---YVGFHYKILK-YYPVDSLOGLNDALCKPRVNGNKAATVPFNEVNTTFS	470	470
Qy	478	RSEYGOVCPDGCVKD-----GIKTHHS-DNDRVRANNEDYKRPWGVKPTNITVLYSGNE	532	532
Db	471	HTTYEDACGWCQAQKEKNGGWMKKEKSCAKKRIYNKE-----NSTDIKILPPEKG	523	523
Qy	533	OGDITQKLEFNCSTNYKDKKNOKWECYKYDEYNINCKLEOMTEIWNDD-----	561	561
Db	524	RKLTLEKLTFCCKQKQKIK--NOIMCHYDNG-----IDDQTDSDNDVLDWGCNLTK	575	575
Qy	582	NPKIISFNFELWYTYILRPTITWMDKLTCTIN-NTTHCIDECNRNLCIFEDRWYKORE	640	640
Db	576	EDKISYAFAPWMMVHDLIDSIKMRBHGKICIMDKGKTCIKGCNKKCICQKWEYQKK	635	635
Qy	641	EEWNSIKKLTFRKKNIOOYSYSNINNLFEGYFVKMDIKDKAKKMELENIKRKKNEF	700	700
Db	636	TEWGIKIDHFKROKDIPKM-----	655	655
Qy	701	SNLENNRPLYENAL--ELLDDHEKATATCKDNNTNEACETSHNATTPCYVRGGTOPT	758	758
Db	656	--THDEFLOTMLMKLLEIIIDTY--GDANE-----	683	683
Qy	759	KNIKIADQYKRSAYEADARNGLHKLGKAHEGYRGGGRKRFKNDICRIAMKHSNRL	818	818
Db	664	--IKRI-----EALLEBA--GVGGIDFALAGLYTKGVAAE--KQTTIDKLLHNEQK--	729	729
Qy	819	GFSGNCPDGGKGTGDIOTRFVVGTEWEVNDREHNRKDHEDVIMPRRRHCTSNLEHLQTD	878	878
Db	730	-----EADKCLTKHTDDTOCPQ-----E	747	747
Qy	879	DHPINGNIYDVLVNSFLGVDLLSAKYEAUKIIRMTKEKNLKGPREVTDPRKHQTTICRA	938	938
Db	748	DRSV-----ARSESXTV-----PSPRADPRATEEV--	772	772
Qy	939	IRYSFADIGDIIRGRDLMEHNGDMVXLOGLHETFGNIHSLKGGKNDKYNDAVKYKTL	998	998
Db	773	-----	772	772
Qy	999	RENNWEANRAKVAEMAKCIDIKYLLKDSGHSOSTOSYCGSDHTPLDDYIPOKLRMWTEMA	1058	1058
Db	773	-----DANAS-----SDDE--DPEEE--	787	787
Qy	1059	EWYCKVQKKEDYLKEKCECKDKNDGSGCTK---ESGNGCKUTEACNEYNIDIIGLME	1115	1115
Db	788	-----EEEBEDGEBEAEBVEQEKTDSEATAVAPSPGGTQ-----	824	824

QY 1116 QWNIISDKYKEHEQAOVSNSGIEASSTAKNHIDRMVLEFLSELYOONGKSNKSGTS 1175
 Db 825 -----DGVAPASOEDDVKCSI-----VDKALKGLDADCTLYKGT--APTS 865
 QY 1176 DESAVIGTNTTYNNGAYLHDTGNFDCCOSQNEFCDEKSDGKNEKAYAPADKQDHDGAC 1235
 Db 866 WKCIPSGNNTTEST-----TKP-----GAA 886
 QY 1236 GCSGSKPTRYOIKTKKAERDECKTVNDILKENDGKQVEDCHPKKNSNCPDMQ 1295
 Db 887 GTPSG-----KDT-----G 895
 QY 1296 NINLVDEPRVCMPPRRKCLVHFL-----ANDN----- 1323
 Db 896 SI-----CVPRRRKLTVGKLHDWAGETTEAKSOETSGGOKTPSGNESSPSEKLPQ 947
 QY 1324 -----EIKKLOSQVNLKAEAFKSAAEFPFVYKSK-----DGEK 1360
 948 GPTPETTKETPESLLHAFVSPPLRFLRFLMFKEDQMKAHAGATGOOTIIGLIDGGG 1007
 1361 NEL-DKELKGRIPPAFLRSMFYTPGDRDLEFG-TDI-----SKGHGSKLKFOIDS--- 1412
 1008 EEPFDLTKGHIIPDFLRKMFYTLGDRDILVGNNDIVYHSGNKEDQIMALQKKE 1067
 QY 1413 --LEKNGDOKSPNGKT-----ROEMWTEHSHIEMALCALV----- 1447
 1068 OILPTSGSSPPRYOTQHSVENPRTKWMENKGIWEGMVCALYNTDTPSGTAPTOI 1127
 QY 1448 -KIGAK-KDFTEN-----YGVNNKFS-----KTTLEEFKAPQFLR 1485
 1128 QERTLRBNSKNPIPOKRYDOVKLDTSDAKTGSPPVSGEKTPLDFTISRPFYR 1187
 QY 1486 WLTEWTDYCYTQKYLKDVOEKCK-----SNDOLKC-----DTECNK-----CEDY 1528
 1188 YLEMGETPCKEKKRLEKKEKRGDRGHEHSGDGYCTRTDADRNKFPDYNCRDC 1247
 QY 1529 VKYMKKKKEM--IPDKYKDEBDRKFRDROHIGVMTDYGYNADYLNKRFKTFASGDR 1586
 1248 HIOCRKRWKIDIKPEYHROEK-----KYQG-----EY-----DK 1278
 QY 1587 PGSASVVRIOLEKQAYYDADKHGCGTFTIENDDKYTNISSKDKCKGLVKEANTGAK 1646
 1279 -----LTKDKSSGGDNCC-----CKDIEKHSAAVF- 1304
 QY 1647 WQNGKPNYNNKLELTDVLFPSRRLRICFHALDGNVTDPEVKDENGILRKLMEVATEG 1706
 1305 -----LKEIK-----HCKNGOTSEKNGQEDOL 1327
 1707 YNLGOYKKEKKEKEKI-KTSDAHKYSTEVPPCSAMKYSFYDLRDIILIGDNLDEKQKTE 1765
 1328 -----NKLDPDKIPOPSPSTYKACP-----YGV- 1353
 QY 1766 ENLKTIENKNGTSVGKSGS--TTGNGSARARFNNENKECYWMANIC-----CYKR 1816
 1354 -----NCGNKRGRKGTNGCTTNE-----PENENKGAASTISILINDSTN 1397
 QY 1817 GRDDGNSNGARSDEDLKCKGVSDDDYPM--GKNDEGTAYOFLWFAEMGEDFCFKH 1874
 1398 GAINGTGT--IDETLKEC-----SDKVAFFGLRROEMTCOK-----KYGVNOC-- 1440
 QY 1875 EKLEKLVGACNDYTCGDNEDKRRKCTDACTOYKKFISEMKPOYEQKIRKYGKNDKIYS 1934
 1441 -----NLTNRYND-TYEDKO-----IVNEFQORLRYF--VHDYINILKRI- 1479
 QY 1935 EHPVADAARAYLDKOLKIKKICENSGDEYECMKDVSTQRLDNGSOMPSLDEPRK 1994
 1480 -----DPCIKKEKODKT--EHKCI-----NGCNI----- 1501
 QY 1995 EVEGKCNQVPRGPRRYRRETPSPRVLISKATASKEAKTAPPTQPKAVENULTTEMA 2054
 1502 -----KGC-----VR----- 1507

QY 2055 QTRBRAAOQTRRRTSTATTESDVTGMKAILSNKPRDSGIEGCPKTYGQYPRMGCI 2114
 Db 1508 -----KWLEI 1512
 QY 2115 VGSKENENGICMPRRKRLCINNIOYLANETENKRDNDKEAFICALIEPFLMKYI 2174
 Db 1513 KG--NEWG-----NIK--KHYNINSNDKET----- 1534
 QY 2175 IENPAENELNGTIPDEFKIMYYTYGDKYD-FRGTDISNDKKIITVNSVTILINEN 2233
 1535 -----IAYNKSIFYDGLDPTDITKKAQKV----- 1560
 QY 2234 NKKODKKDEELKIFEMENKKEFIWEGMITYGLYHLTDE-NEKEKIRNTOYONDYKLT 2292
 Db 1561 -----EDEKERRKIW-----GCTGH--DECSERKEENK--NFTNLJ 1594
 QY 2293 PSLEFPVKRRQFLRWFTMAEFECNRKRBOLLLEAGCKRYEBCNGSNDCTOCACBYT 2352
 Db 1595 SELQDKITSCQ-----NKH-----NPNKGT-----AC-- 1616
 QY 2353 YONFIKWKTEYERQREKFKDKDKKDYKDYPTERDIEKATCAHEYLANKLKELCGNKD 2412
 1617 ----- 1616
 QY 2413 CSCMQRPSSOLPPTQOSQSSDANDPESLDYVPEEFNKECEBELSKKSGMHTKITEP 2472
 1617 -----DFPSPTEPTEPDLDDTPDPLDDQHTQEPQFCP----- 1653
 QY 2473 KIPMNCYKAAVYLSKAEENMDITLKEK-----PIPIESRK-EKESK 2514
 Db 1654 PPEPTCYEKLAKELEYAEKGINNELKNGKDFNGKCNVKKNGAVIGEESCKFEQTYE 1713
 2515 NSWTNNPCDPKRPYAPDKYIGRRNPENRENEFRKYDYMCKYKNSKFOE--KRRVCP 2573
 Db 1714 NSVNNIN-----NCKDONERFKIGQW--NFKYGTIRKDLCP 1752
 QY 2574 PREHMCRLNDELKIERLKDSNYLKMVRRTARNEGIDITIKNFENSEGCMANPIDCTMK 2633
 1753 PREHMCRLDLSMIGRTTISDSSALLKIOEAKSEHDDIIRKLEQNSCDEHRICDAMK 1812
 QY 2634 YSFADGDIYRGDMLRIGYLPPEVEIKLYKVEYIYGGKRN--KN--KARNYNDVQTF 2689
 1813 YSFADGDIIRGDLNKNKSKQGLQRLLEYAFLINYNKQNDNKKYKORPYLDV-- 1869
 QY 2690 RSAMWDANRRDIWKAMCKAPEDAKLEFRKGRMD-----GFERITLLQDKGHRDDPEVD 2744
 1870 RSDWMDANRRKHINNAMCNPDAKFLKKNPNDTSSSSSKGIMTTHSNGYKPEPDYD 1929
 QY 2745 YIPQRFMMTEWSBYCKALMBELKFKKSCDHC-KTSDRCNDYDENKCEOCKTRCOEY 2803
 1930 YIPQRFMMTEWSBYCKALMBELKFKKSCDHC-KTSDRCNDYDENKCEOCKTRCOEY 1989
 QY 2804 KNEFLKWKSLFDIOSNKKYKLEYROPY--TKISTYDHQONVOYOKLTKFSKC-SVESF 2858
 1990 KKLHNKLGFD-----KYKITYNE-IYNNKOSKINSNEYKFLKLEK--KDKCKELNSS 2041
 QY 2859 SEYLHEKSKLANKFENENDSSNIRTYAFETEPKSYEACSTLPSKNPLNDPTDO--N 2916
 2042 DKCIDEAFTGCTYKFS--NSENKHNHNNVAFKNPREYKAKACKDAP--DPLDNCPEKSDATY 2098
 QY 2917 KDCKELQFTFPCSKNDYDNNDLNNVAYLVNSSDDKGVYLPERRRHLCRPTIAYNRY 2976
 2099 EKACNTLLPRLCKESKTFNNDDSDMTSFVQTSBPDGTGLVPPRRRQICLKNITTY- KLR 2157
 QY 2977 KGDK-EIKKIKLLTSAFSQCQLLGOKKSEELCFEAMKSYADYSIITCTMDMTSLS 3035
 2158 SIERIDFKALMTSAENKELCEIKKRDVTLQAMKSYFDYGTIVAGTDLISTAPL 2217
 QY 3036 EIKIKIIF-----ETSNEAENRKTWMENNRROIWMHMLGCIATSKYVLIDGQWCOLPK 3089
 2218 DKLKTNLVNLKGDGTNEIEDRGKWWTEBNTRYWAMHMLGCIYAAAGKI--EBRDCSLP- 2274
 QY 3090 DEETNOFLRWLITMAKOACEKKHVSLSLTKCPRSNEDNEFASELLROPGCO--NDIR 3146

Db 2275 DONTHOFLEFREWSEHFCARQKLEFNEVREC--ASAOCIEYGTIDPVCCEACTQYR 2332
QY 3147 KYISLNLILIKTMEMLNKKYKOLKODSSGNNDKPSK--ENVQSYIKSKDSCALMLINDIN 3205
Db 2333 DYIRKIQ-----ERLNLNYQNTNFNEKKALEYTAPEYFNDK---CNDKNCLS 2379
QY 3206 EYVGTGKNNNEKPEKLYKLYVEDETHKNNHVDGNKEEQTVKPKALYFETPH 3265
Db 2380 KYI-----DIEKKMKMYDSF---DDNDLKNKCICQIKPKRP---PKVKKPEEH 2424
QY 3266 VDSFYQAPLSTHVAQYD-----KNDLKSSIVYISAGLIALHMKKKKSSV- 3318
Db 2425 TPSSQDTPPLPKPPKDDLPAPAEPPNRDILETIPFGIALAGSIAFLKTKKSSVG 2484
QY 3319 DLRLIINPOGEGMPLLESKNRPIYRSQPKTYIMEGDT---SGDEDYMDLSS 3375
Db 2485 NLFQILHPRKSDYDIPKLSPNRPIPTSGYKGRITYLEGDSGDSGTIDH----- 2538
QY 3376 SDIT-SESEYEELDINDIYVPSPKYKILIEVLEPSKDDIPSD--DPS-----ND 3425
2539 SDITSESEYEELDINDIYVPSPKYKILIEVLEPSGNNNTTASGKNTPSDTQNDIQND 2598
QY 3426 TPRTNFIIDEMNELKHDFEYSQYL---PNTPENNNYKSADIPMNTPENTLISDNPEKPP 3482
Db 2599 GIPSKITDMEWMLTKDEFISNMLQNEPNER--NMLGYVNDNTHPTT--SRHNVKEKPP 2655
QY 3483 IISIHRODLYTGKEISYINIMSTNTNNDIPMARNDSTYRGIDLINDSL 3530
Db 2656 IMSIHRODLYSGEYSYVNM---VANDIPISARNGNYSIDLINDSL 2700

RESULT 10
Q9U4A2
ID Q9U4A2 PRELIMINARY; PRT: 1685 AA.
AC Q9U4A2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAY-2002 (Tremblrel. 20, Last annotation update)
DE Variant surface protein PfEMP1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5833;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE-20144115; PubMed-10677532;
Smith J.D., Grais A.G., Kriek N., Hudson-Taylor D., Kyes S., Fagen T.,
Pinches R., Baruch D.I., Newbold C.I., Miller L.H.;
"Identification of a Plasmodium falciparum intercellular adhesion
molecule-1 binding domain: A parasite adhesion trait implicated in
cerebral malaria";
PL Proc. Natl. Acad. Sci. U.S.A. 97:1766-1771(2000).
DR EML; AFI93424; AAF18980.1--
DR InterPro; IPR00454; Eub_Atpase_Csub.
DR InterPro; IPR004258; PfEMP.
DR Pfam; PF03011; PfEMP_1.
DR PROSITE; PS00605; Atpase_C_1.
FT NON TER 1685 1685
SQ SEQUENCE 1685 AA: 193671 MW: D1PD42666B0551E CRC64;

Query Match 15.1%; Score 2932; DB 5; Length 1685;
Best Local Similarity 38.6%; Pred. No. 1.2e-133;
Matches 650; Conservative 265; Mismatches 619; Indels 152; Gaps 51;

QY 33 KSARVLEERYAKNIRHPKRYKAHEVDS-LKGDLYTAERGGSPSPVNNKNNYTPPCND 91
Db 8 KSARVLEERYAKNIRHPKRYKAHEVDS-LKGDLYTAERGGSPSPVNNKNNYTPPCND 91
Db 8 KSARVLEERYAKNIRHPKRYKAHEVDS-LKGDLYTAERGGSPSPVNNKNNYTPPCND 91
QY 92 KHEHTNLRYDDVNLRHPCGHEORNFDEDESEC-GNIRYKKNKNDAL-ACAPPRRHH 149
Db 61 HTRDINVTIGQ-GREYFCGQVRFSEDRSQCMSNRIRG--SENDVYGACAEYRLRHV 117

QY 150 CDKNLEALNDINTQNIHDLGNVLTAKTEGSIYNNHP--HKGISDACTALARSFADIG 207
Db 118 CDKNLEALNDINTQNIHDLGNVLTAKTEGSIYNNHP--HKGISDACTALARSFADIG 207
QY 208 DIYRGIDMEKPPNVH--DKVETGLREYFKIHDMGHEVDYNDNPGSSGYTYLREANMAY 265
Db 178 DIYRGIDMEKPPNVH--DKVETGLREYFKIHDMGHEVDYNDNPGSSGYTYLREANMAY 265
QY 266 NRNKWEAITCDASYSKGYFMQSESNTPLSPNPKGKHOKGVPNTDYPQYLRWFEDWG 325
Db 237 NRDKWKAITCTAPQAYFRAACGGEHTMTYGQRCVNGEPTNFEDYIPQYLRWFEDWG 296
QY 326 EECRCRNKILKLVDSQNDKRELKCSHNGDCTTTTKKQITLILDKCTDCSTCKYF 385
Db 297 EDCRCRNKILKLVDSQNDKRELKCSHNGDCTTTTKKQITLILDKCTDCSTCKYF 385
QY 386 EYWLGNQOBAFKQKREYKELQSYLSNDNKNVNNINSEYKQFYKELKETOYATNDFL 445
Db 357 YKMGNKREKPEKQKCKENEYKNNKNOTOSSISVNDMYDHFYSLEKE-KYSSIHFT 415
QY 446 NLLNGKCYKGLPBEKIDITFNSADKGIYRSYCOVCDPCGY--KCDGIKYTHKSDN 503
Db 416 NLLNGKCYKGLPBEKIDITFNSADKGIYRSYCOVCDPCGY--KCDGIKYTHKSDN 503
QY 504 DREYVNNEDYKPPMGVKNPTNITVLYSGNDOGDITQKLEPNCSSTNYKDKN-NQKMECY 562
Db 474 DCACHKNPYTPNNQVKKPELTVSSGGLDLEKLEPKFTSDNNKDFSLYEBWKY 533
QY 563 KDNINROCKLEONTEINNDNPKI-ISPNNFELWYTLRLTIKNDLKTICINNTHTC 621
Db 534 EHDGNEACILKKEKSKRAEIQKSFEDFTYVAHMLKQSEDMKTLSLCINKKQDLC 593
QY 622 IDEGNRLCFLDPMWYKQKEEEMNSIKKLFTKKNIQOSYSINNNLPFGYFFKYMADLK 681
Db 594 INNCNRKCYEKTMEKEEEMEKIKVHEKQKPEGHAYVILGYELDFIKITITNAYG 653
QY 682 DEAKKELMENIKRKNFSLNENRDYLENAELLDHLEKATFCQNDNTNA-C--- 737
Db 654 EKEVQYIAELLRNKTKGSPDQATQ--KTIIDELDHELEBAHECRENNPEKNSKE 710
QY 738 -----ESHNAATTPCYKPPGTOPTNYKIEIAQYFRSAVEEARN-----R 779
Db 711 PHDDDEDETHY---NPGCKTDG---TVRAQIKAKKFORDAKQMKNTTRDGTGRK 763
QY 780 GLH-KLKGAHEGYIKRGRKDPK-DNLCHIMIKHSNRNLFSGNGCDGCGDGTG-- 835
Db 764 GAHNSIVGDISKAFKNGGOSDLKDKIDINTSHNSDSNGGCGCTGK---DGNQGG 820
QY 836 TRFYVGTWEVDPPEHMKRDEHDVIMPRRRHICTSNLEHLOTDPHPLNGNIVDLVNSF 895
Db 821 DRMKIGTPWSKVGD-KTYSDVYLPPRQHMCTSNLEFETKQTPLDGKFGVDKINHSF 879
QY 896 LGDVLTSKYVANKIIRMYEKNLKGKPEVTPKHQTTICRLAIRYFADIGILIRGL 955
Db 880 LGDVLTSKYVANKIIRMYEKNLKGKPEVTPKHQTTICRLAIRYFADIGILIRGL 955
QY 956 WERNGNVVLQGLETFVGNHSLKSGKNDYNDAPKYLKIRENWEANRAKVEAMK 1015
Db 937 WDKDEGSKKNDVILKIFGRIKQELPREIKQKKNPDGKITQKRWMEANRQVRAK 996
QY 1016 CDIKYLDKSGHSTOSYSGSDHTPLDYIPQKLRMTTEMAEWYCKVQKKEKLEK 1075
Db 997 CAL-----DQGSIEK---NGIPLDYIPQKLRMTTEMAEWYCKVQKKEKLEK 1044
QY 1076 CKECKDNDGOGCTKCTACNEYNIIIGLMEKNIIISDKYKELHDAQMSY 1135
Db 1045 CSQCKTMD--KKCTNS--DCCNTCTEACTAYNNKINTWQOMADISDKYKELHDAQMSY 1101
QY 1136 SNGSIEASSTAKNHNDRNVEFLSELYOONGK-SNKSQSDSAVY-----GTWTTYE 1188
Db 1102 ANGGPHASSGSDVEKDKPVVNFLELYKQNGKISIPSDIHPQRYKRAAPSNSVTYVS 1161

QY 1189 NVGAYLHDTGNPDCCOSQNEFCDEKSDKD-NKRYAFROKPDHDCAGCKSGSKPTRVQ 1247
 DB 1162 TAAGYIHQEAHIDCKNQVNECEKKGKGGNNKRYAFHPEYDHKACACDGRNPVKV- 1220
 QY 1248 IKTKKAEKDECTKYNDILKENDGKQYEDCHPKNSNGYDPMOCGNINLYEDPVCA 1307
 DB 1221 -----LEDECKAVOKILISFOIEKNNIHC--KTTERA--KWCENTKLEDEGVC 1267
 QY 1308 PPRRQKLYHFLANDNEIKKQSOVNLKEAFIKSAAEFFSWYXXSKDGEDELDEL 1367
 DB 1268 PPRRQKLYHFLANDNEIKKQSOVNLKEAFIKSAAEFFSWYXXSKDGEDELDEL 1321
 QY 1368 KEGKIPAPLRSMTYFGYRDLFTGTDISK--GHGSGSLKQOISLFPNGQSKNGK 1425
 DB 1322 HRDMPPEFRSMFTYFGYRDLFTGTDISK--GHGSGSLKQOISLFPNGQSKNGK 1381
 QY 1426 ---TROEMTEHSEIWEALCAL--VKIGAKKD-----FTENYGNVNFSDKSTLEE 1476
 DB 1382 KVLREGMKEVSLTKMGKLCALSTNTEKKNDEGVRYTIAMTYKN--NDIKETLEE 1438
 QY 1477 FAKRQFLMLTMYDYCYTRQKYLKDVQEKSS-----NDOLKCDTE-----CNKR 1524
 DB 1439 FAKRQFLMLTMYDYCYTRQKYLKDVQEKSS-----NDOLKCDTE-----CNKR 1498
 QY 1525 CEDYVYVYMK--KKEWIPQKYYKDEKDKKREDQHGVMYTDYTGYNATYILNRKFTASC 1583
 DB 1499 CERKYYKMEKRYKHYSSQKKKFKQLYKNSATYNN--GLAVKE--ANSEYKKNDPEVTEA 1552
 QY 1584 GDRKPSASVVO---RNIOLEKQAYYADADKHCCTKFIENDD-----KYTNISSKDKCKG 1635
 DB 1553 NSAKHARDYIKTQLENNICNGMYKNCYTCANTSSSTNSBPASLDITPSEYKDKCNC 1612
 QY 1636 LYKCAN 1641
 DB 1613 VPDECS 1618

RESULT 11
 ID 087560 PRELIMINARY; PRT: 2658 AA.
 AC 087560;
 DT 01-JUN-2002 (TREMELREL. 21, Created)
 DT 01-JUN-2002 (TREMELREL. 21, Last sequence update)
 DT 01-JUN-2002 (TREMELREL. 21, Last annotation update)
 DE Erythrocyte membrane protein-1 (Fragment).
 GN VAR.
 OS Plasmodium falciparum.
 NC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=5833;
 (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-JDP8;
 RA Chatopadhyay R., Pillai C.R., Chitnis C.;
 RT Identification of a domain responsible for binding to intercellular
 adhesion molecule-1 from a Plasmodium falciparum field isolate.
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBD databases.
 DR EMBL: AY028643; AAK49742.1; -
 FT NON_TER 2658 2658
 SQ SEQUENCE 2658 AA; 301326 MW; AEE33BC9167E7B7 CRC64;

Query Match 14.8%; Score 2869; DB 5; Length 2658;
 Best Local Similarity 27.6%; Pred. No. 2,3e-130;
 Matches 869; Conservative 366; Mismatches 906; Indels 1012; Gaps 122;

QY 31 SHSARVLERYAKNIRHPSRYAKENHVSLSKD-----LTAKFRGSPSTPVNKHYY 84
 DB 11 SGTAEELDRIGEV-----YKAHDAKAGSMGLKGLTSLQAIFFKAPKQKSTSEB--- 62
 QY 85 PYPCNDLKHKEHTNR--YDVNLRHPCHGREGNRPDEDESECG--NKIRNYKKRNDALAC 141
 DB 63 --PCDLNHEHTYVTSYGDKEN--PCKDRPEVFSYTEGAECDSKIRGSNSNKGDC--AC 116
 QY 142 APRRRRMCKNLEAL--NDINTQNHDLGCVLTVAKYBESF-----VNNHPIKGT 192

DB 117 APRRLHLCDQHLHKKHDTTR--HNLLADVEAAFAEASLEKRGQYOLNN----- 168
 QY 193 SDA-----CTALASFADIGDIVGIDMFKPNVDK-----VEGLREVFKKID-----G 238
 DB 169 SDVININCTELARSPADIGIVGKDLFLGNTYESAORILLENKLEIKETKIHSDVTKG 228
 QY 239 MEDEVKNDYVPGSGNYYKLEAMVNVNRKWEATITGASYSGYFMOSESMTPLSNP 298
 DB 229 TNGELKRYEKDQ--NEYQLREDMTYANRETYVATIGAGTSDBKFRNCSNKGATSG 287
 QY 299 KCGHKGQVPTNLDYVQYLRWDEWEEECRKNIKKRVKOSCR--NDEKLYCSHG 356
 DB 288 KCRKNDQVPTFYDYVQYLRWDEWEEECRKNIKKRVKOSCR--NDEKLYCSHG 346
 QY 357 HDCTITMKGIHLDKCTCSTCKVFWMLGNQOAKKQKKEKEKELIOSYLSNDK 416
 DB 347 YDEKTAASGHDFEDDYDCQSCAPFDYLDNQLEFLKORNTYKEMOKY--TNGTK 405
 QY 417 FVNNINSE-----YKQYELKETQYATNTDFLNLMEGKYCK-----G 456
 DB 406 ---NSKRKKPGAGSNYDGEKRFYNI--EGYNNVDFLNDLDETCKKNNEIEG 459
 QY 457 GLGEKDIPTNSADD--KGIFRSEYCOVCPDCCGYKCGITV--TKSNDKBRVANNED 512
 DB 460 GQINFKNVNSGKNSDDDSKTKTCRTTYCOACFWCGAEEDNSGKWKANADACGRKD 519
 QY 513 YKPPMGVPTNIVYLSGNQGDITOKLEFNCNST-----NYK----- 551
 DB 520 YDPE---KTTIELIGDTRKSDMYQKFKCTSATATGAPDTATGEGNGKAGASCKNG 576
 QY 552 DKNNQWECYKQDEN-----INRCKLEONTIENDNPKIISFHFELWYTLARDT 603
 DB 577 DNITEWTQYKRRKNGDYVKKIDNFCVLQDQK--HTKEQKVSYNVFEKMYVDMILHS 635
 QY 604 IKMNDKLCQCIINT--THCID--EONNCICLPDRMYQKKEENSTIKLE--TKKNIQOS 659
 DB 636 IKMNEELRSCINNAKSONCKNNKCNDSGCFEWEIQKREKDKIKDHRKTDFGIPG 695
 QY 660 YYSNINNLFEGYFFKYVDKIDKDEAKKELMENIKRKNFNSLENNROYLENA----- 713
 DB 696 PLOQFD-----YDFVLKFTVDK-----KELLQNKOTHADAKIGIREKMLEQAGVDSG 745
 QY 714 -----IBLLDLHKEATYICKDNNTNACETSHNATTPPCVPRGQTPT 758
 DB 746 SGGGGNGAKGKHNTKIDKFLQEEQFAETCK--QTONCPKPKPYANPCY---GNNTYD 800
 QY 759 KNIKETAOYKRSAYEARN--RGLHRLKGAHGIYKRGGRDRDKNLCRIMKHSNRN 817
 DB 801 ALAEKVAQILQGEAQVQLINHRSSSLKGDPEQGHYNGNKNKSVLKD--VCKITNOYSNAG 859
 QY 818 LGRSNGPDCGKGTGDTGTFQTFVYVGTWEVDPEHRRKDEHVIAPRRRHICTSNLEHOT 877
 DB 860 KNSNDPCNCKN-----ENRPNIGEKKNKGEVAKMSTHS--YAPRREHCTSNLEHLK 913
 QY 878 DQHPNLGNIVDVLNNSFLGDVLLSAKYEANKIIRMYKKNLKGPEVTPDKHQTTCR 937
 DB 914 DKGRFQVDPDNKATSHLADLILAAKKEAIKELYEKKNDGSKNAKGNLNEKTVCR 973
 QY 938 AIRYSPADIGDIIRGRDLMERNNGDMVLOGLHLEVPENHLSLKGNKD--KYNDAKPY 995
 DB 974 AIRSPADIGDIIRGRDLMERNNGDMVLOGLHLEVPENHLSLKGNKD--KYNDAKPY 1027
 QY 996 IKTRENNMEANRAKVMAMCDIKYLLDKSGHOSQSSYSGYSDHTPLDYITPOKLEHMT 1055
 DB 1028 IELRKQWMEANRAKVMAMCAKRDIDPMK-----CNGPIDIEDYITPOKLEHMT 1075
 QY 1056 EWAWEYCKVOKKEYDKLKECKECKDKDNQOGCTKRESGTG--CTKCTEACNEVNDIIGLWK 1114
 DB 1076 EWAWEYCKVOKKEYDKLKECKECKDKDNQOGCTKRESGTG--CTKCTEACNEVNDIIGLWK 1131
 QY 1115 EOWNIISDKKELHEQAOQMSVNSGIFASSTANHNIDRNIIEFLSELYQO-----NGK 1168


```

Db 1132 QQNDKMLKTIILYANKTTSTNAGRTVLCDSADPY-QQMLDEFKLOKEIKNSALNRTK 1190
QY 1169 SNKSGTDESAGVIGTNTTYENGVAYLHDTGNFDCOSQNEFCDEKSDGKNENKAYAFDRK 1228
Db 1191 RSJDTGNNDPIF---TSPYSSAGVYIHQELPHYQOVQAKFCFTNG---NKEKAYARHOP 1244
QY 1229 QDDHAGCGSSGSKPT-----RQIOTK-----KKAERK 1258
Db 1245 HDTDEACACRPSTPESPANLPLADOGKVEEAEEREEEDVEDDDDDDEBEEREEB 1304
QY 1259 TECKTVNDILKENDKGOVEDCHPK-KNSN-GYPMOCGNINLVEDP-RVCMPPRROKIC 1315
Db 1305 PCKTAVNDILSTDRTKQVQOGCAKIKTNINESTPDMTCVANSKENNEYCGMPPRROKIC 1364
QY 1316 VHTL-----ANDNEIKLOQOVNLKEAFYKSAAEFFESWYYSKDGEGNEID-KEIKEG 1370
Db 1365 LYYLKEJGEHNDDEOK-----FKDAFIKYAAEFELSMQYKSK---NSMDIKLOSG 1413
QY 1371 KIPPAFLRSMEYFEFGYRDLFTGDTISKGBESKLEQIDSLFKNGDQSPNGKTRQEM 1430
Db 1414 EIPBEELRSMTFTYGDYRDLCLNTDLSKKEGVDYDAKGIADAFNKYTDTN-----RTKW 1468
QY 1431 WTEHSEIWEAMCAL---VKIGAKKDPFTENYGNVNFVSKSTYLEEFARPOFLRWL 1487
Db 1469 WDTNGBEIWEAMCALHTGVTMTDNKRKIKTDYSYKELQSKVNTTLEKFAERPOFLRWL 1528
QY 1488 TEMYDYCTRQKYLKDVQEK-----SNDOLKCD-TECNCKCEDYVAYMKKKEM 1538
Db 1529 IEMGDFECBQKKYKELKCNKCCNNGNVTSDCKTCVCEOKCEYKGF---TEW 1585
QY 1539 IPDCKYKDERDKRRDRHIGVMTDYGTATDYLNRKFTASGDKRGSAVQORNQ 1598
Db 1586 --QENMNK---QNKKEITLYQKTSRSTISSDPIEKL----- 1621
QY 1599 LLEKQAYYADAHCGCTKT-ENDDKYTNISSKDCCKGLVEANTGALMOKNGPNYNN 1657
Db 1622 -----LKTINELKDPYGNSTYKRAAGYTK----- 1647
QY 1658 LKEITEDVLPFRRLRICHALDNGNTDEPVDENGLRKRLMEVAETGYNLGOYKKEK 1717
Db 1648 ----- 1650
QY 1718 EKEKIKTSDAHKSYEVPKCSAMKSYFDLNDIILGIDNLEDEKOTEENLKIFPNKGT 1777
Db 1651 -----IEGCNVSQONFDP----- 1663
QY 1778 SVGKGSDDTGTGNGSTARKFENNEKECVWNAITGIRGRDDNGSGSARSDEDLKCG 1837
Db 1664 -----ENK-----NGN----- 1670
QY 1838 SVPSDDDYPMGNRDEGTAYOFLRWFALMGEDFCHEKKELEKLVGACNDYTCGN----- 1893
Db 1671 -----DEKAF-----KEPKETEK---ACK---CNENTPPS 1696
QY 1894 -----EDKRRKCTDAGTQYKFISEMPPQYKQIKKYGENKDKLYSEHPVAKDAEDARE 1947
Db 1697 PPELPGPAPATDYSVYCEVTSALT----- 1721
QY 1948 YLDKOLKIKENKSGD---CEYKCKMDVSTQRLDNGSNOMPASIDDEPKVEKGCQY 2004
Db 1722 ---IENLAKACPTKYGSKAPYTMKCP---SGEKSGAGATERSRADGAPSSSSEGSJCV 1777
QY 2005 PRGPVRVRETPSPRVLISKATASKKEAKTAPPTKOPKRENTLTERAOTRTTBRAAQ 2064
Db 1778 ---PPRRRLTYVGLQEAKEKTSQAEGSSATQOEGVYKNN-KDAGSSSSSSNSQYOT 1833
QY 2065 TRKRTSTTTTESDVGIVYKAILSNKPDSSKGIQCNPKTYGQYPKWGCIVGKSKENENG 2124
Db 1834 TLNASTSTSTES----- 1846
QY 2125 ICMPPRRKLCINNIOYLVNTEKRNDRDIKAFIKALIFTOPLMKY----- 2173
Db 1847 -----SOLLQAFIQAIALETFLMHRYYKKEKERR 1880

```

```

QY 2174 -----IENPAENELONGTIPDEKRIIMTYGYDKMFEGT 2211
Db 1881 RKKKKEHSTGTLEGLSVDEEEQPPQEBLOSGKIPYGYLRQMYLADYKILYSG 1940
QY 2212 DISNDRKII-----TVNSVTTIENENKKQ-----DKKDEELRKIFW 2251
Db 1941 SMDNLKHIYLAESGTREKQDMQKQDKIKKTLNDNNOESHPPSPSKNSVTTPTQYTW 2000
QY 2252 EKNKFTWGMITGLYTLTD-----ENEKIRIDNYQYNDMTKL 2291
Db 2001 NHAAPSIWGMIRALTHKSDATYIAGGKIEQNTOLKNAALLDPTKKNKPDYKY-DYVKL 2059
QY 2292 -----TPSEFVYKRPQFLRFTWAEFCKRREOLKLEAGK----- 2331
Db 2060 DENSGTSPRPAVNNAPRKLTEVERPPYFLWLEWGSFCRKHOKHLEIIRVDCRGE 2119
QY 2332 EYECNGSN-----DEKTOCAEACVATYQNFITKWKTEYEROREKPKDKD 2376
Db 2120 DKHCSGYGENDNLADPSIRPDLNCPGCAKHCSYKWKIORKKIEFTQODNAYNN--- 2176
QY 2377 GKKYDYSTERDIKATCAHAYLMLKLELCGNKDCSCMOQPSOLPRTYQOSOSDAN 2436
Db 2177 -----QKVCERK-----SKGGNGVCCKLE----- 2197
QY 2437 DWPESLDYVPEEFNKCCEPELSKSGSMITHKITEPKIPMNCVKAAYLSKE---AENN 2493
Db 2198 -----ITLEKFLPIESTEKESKNSWTNNPCDPKPPAPDPKYIGRRNFCENREBNR 2548
QY 2494 MD-----ITLEKFLPIESTEKESKNSWTNNPCDPKPPAPDPKYIGRRNFCENREBNR 2548
Db 2215 KDNGDGTINFEPEVDYTK-----PADNCK-----PC-----SE 2242
QY 2549 FKVDYEMCYKNSKFEYOKRRVCPPRREHMCNLNDEIKTERKD---SNYLLKAY--- 2602
Db 2243 FKID---CTAKCTGDEBKRC-----NANNEYVIRATDITKDKNGENENIMVYSD 2290
QY 2603 -----RRTANEGIDITKNNSE-----NCCANMPCIDTMYKSYFADIGD-----IV 2643
Db 2291 TSKKGDODDLHYCHDAG--IFKGIKRDWICGNVCGVN-YCKPYKVN-GQSGDNOQIITI 2346
QY 2644 RGTMLRIGYLV---PVEILYK-----VEYIYK---WRN----- 2675
Db 2347 KAFKRIWATFLEDYNNKIKKLSCTSSDAPICIGCALFWLKKTKTEWKNLNYLQ 2406
QY 2676 -KNGRKNYNDVQTFPSAMDPANRDKIWKAM-TCKAPEDAKLF-----RGRMDGFE 2725
Db 2407 YENSSDKSFLVKTILBEFD-RRFQNAIKPKCELEKFEESFGLNGADNSKSKDKGE 2464
QY 2726 R---ITLIQ-----DYCGHKD-----DPPV---DDYIPORFRMTWMSSEYYC 2761
Db 2465 RDLVLCITIEKLEKAEKCAENDADQNGVQCTQYTTDPLEDEDL----- 2509
QY 2762 KALMEELEKFKSCDHCKTSDRCKNDYDENKCE 2794
Db 2510 ---LLEEBENTYKAFEGCEMKETKDEBECK 2540

```

```

RESULT 12
08T6LO
ID 08T6LO PRELIMINARY; PRT: 510 AA.
AC 08T6LO:
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=5833;
RN [1]
SEQUENCE FROM N.A.
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;

```

* Identification of a conserved Plasmodium falciparum var gene
RT implicated in malaria in pregnancy. "

RT J. Infect. Dis. 0:0-0(2002).

DR EMBL: AF458886; AAL84269.1; "

FT NON_TER 1 1

FT NON_TER 510 510

SEQUENCE 510 AA: 60713 MW: E3CF55DC04601099 CRC64;

Query Match 14.5%; Score 2817; DB 5; Length 510;

Best Local Similarity 98.2%; Pred. No. 9.8e-129;

Matches 501; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 233 KTHHGDMEVYKNDYNDGSGNTYKLRAMWVNNRKNVWALITCDASYSKGYFMQSSNT 292
DB 1 KTHHGDMEVYKNDYNDGSGNTYKLRAMWVNNRKNVWALITCDASYSKGYFMQSSNT 60
QY 293 PLTSNPKCGHKGKVPNLNDYVPOYLRFWDEWGEBCFRKNIKLKVKDCSRNDKRLYC 352
DB 61 PLTSNPKCGHKGKVPNLNDYVPOYLRFWDEWGEBCFRKNIKLKVKDCSRNDKRLYC 120
QY 353 SHNGHDCVTTIMKGLIHLNDKCTDCTPKCVFVWLGNOQAFKKOKERYEKI0SYLS 412
DB 121 SHNGHDCVTTIMKGLIHLNDKCTDCTPKCVFVWLGNOQAFKKOKERYEKI0SYLS 180
QY 413 NDKKFNANNSEYKQFYEKLETOYATNTPLNLNEGKCKGGLPGEODITFTNSAD 472
DB 181 NDKKFNANNSEYKQFYEKLETOYATNTPLNLNEGKCKGGLPGEODITFTNSAD 240
QY 473 KGLFYSEYQVCPDGCYKCDGIRYTHKSDNDRVNNEDYKPPMGVKNPTNITVLSGNE 532
DB 241 KGLFYSEYQVCPDGCYKCDGIRYTHKSDNDRVNNEDYKPPMGVKNPTNITVLSGNE 300
QY 533 QGDITOKLENCSSSTNYKDNKNQKWCYKRDENINRCKLEONTLEINNDPKIISFHNFF 592
DB 301 QGDITOKLENCSSSTNYKDNKNQKWCYKRDENINRCKLEONTLEINNDPKIISFHNFF 360
QY 593 ELWVYTYLLRDTIKWMDKLCINNTTHHCIDECNRNCLCFDRVYKQKEEEMNSIKKLEFT 652
DB 361 ELWVYTYLLRDTIKWMDKLCINNTTHHCIDECNRNCLCFDRVYKQKEEEMNSIKKLEFT 420
QY 653 KKNIOGSYSNNINLEGEYFFKWMKLDKDEAKKWELENIKKRKNFESULENNRODLEN 712
DB 421 KKNIOGSYSNNINLEGEYFFKWMKLDKDEAKKWELENIKKRKNFESULENNRODLEN 480
QY 713 AIELLDHLKETATICKDNNTNEACETSHN 742
DB 481 AIELLDHLKETATICKDNNTNEACETSHN 510

SULT 13
Q26033
ID 026033 PRELIMINARY; PRT; 2664 AA.
AC 026033;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Variant-specific surface protein.
GN VAR-2.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MEDLINE=95330813; Pubmed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
Peterson D.S., Ravetch J.A., Wellem T.E.,
"The large diverse gene family var encodes proteins involved in
cytoadherence and antigenic variation of Plasmodium falciparum-
infected erythrocytes.";
RL Cell 82:89-100(1995).
DR EMBL: I40609; AAA75398.1; -
InterPro: IPR004258; PFEMP.

DR Pfam: PF03011; PFEMP; 2.
SEQUENCE 2664 AA: 302410 MW: 6EA2468511703091 CRC64;

Query Match 14.0%; Score 2714.5; DB 5; Length 2664;

Best Local Similarity 23.9%; Pred. No. 7.1e-123;

Matches 903; Conservative 395; Mismatches 879; Indels 1605; Gaps 123;

QY 33 KSARNVLEARYAKNIRHPSYAKEHVS-----LKGLTKAEFRGSPSTPVNKNHNYTP 85
DB 17 RSKAHILDSIGKKV-----YDKVHGDAIQPSNGKLTSLAIFEKAPBEKQTSSED----- 67
QY 86 YPCNLDHKEHTNLR--YDVNVLKHPCHGROONPDEBESDCG-NKIRATKRAKDAIACA 142
DB 68 -PCDNLNHYHTTYSQDKEN---PCKDREVRVSYEGAECDKSKIRGSNSKDG-ACA 122
QY 143 PPRRHMCNCKNLAL-NDITONIHDLGVNLTAKYEGESI-----VNNHPKQTS 193
DB 123 PFRHLCDQHLHKKIKDKTR--HNLADYCEAKFEASLEKRYQYQLN-----S 174
QY 194 DA---CTALARSFADIGDIVRGIDMKPN--VHDKYETGLREYFKKIHGDMED-EVKND 246
DB 175 DVNINICTELARSPADIGDIVRGIDLYRGNDKEKRLLEENLRKIFKKIYDNLDAVQEH 234
QY 247 YNPD--GSGNYTKLRAMWVNNRKNVWALITCDASYSKGYFMQSSNTPLTSNPKCGHKQ 304
DB 235 YKDDKGTKNYKLRAMWVNNRKNVWALITCGAA--GGLFYRQTCGTGTWTEKRCPI 292
QY 305 KGYPTNLDYVPOYLRFWDEWGEBCFRKNIKLKVKDCSRNDKRLYC3SHNGHDCVTTIW 364
DB 293 NDPTTYDYVPOYLRFWDEWGEBCFRKNIKLKVKDCSRNDKRLYC3SHNGHDCVTTIW 352
QY 365 KGLIHLNDKCTDCTPKCVFVWLGNOQAFKKOKERYEKI0SYLSNDKNFNANNINSE 424
DB 353 VIGHVHVGSCSKSCYCRKRWIDNQKEFLKQKKCEMELSKKKQSTRYNYEG 412
QY 425 YKQFYFKLETOYATNTPLNLNEGKCK-----GGLPGEODITFTNS 469
DB 413 YKQFYFKLETOYATNTPLNLNEGKCK-----GGLPGEODITFTNS 472
QY 470 ADDKGIYFSEYQVCPDGCYKCDGIRYTHKSDNDRVNNEDYKPPMGVKNPTNITVLS 529
DB 473 NNSKRTSHSQYCEECGCGVELIGNEM--KEKKGCKGKGRYNDPKTHNYIVPLSF 530
QY 530 GNEGODITOKLENCSSSTNYKDNKNQKWCYKRDENINRCKLEONTLEINNDPKIISF 588
DB 531 GDEKKEIKETIEQCAENSDDSKLEQWMCYCYGDKRYEYCTLENRKNSSEDPEDIKTE 590
QY 589 HNFELWVYTYLLRDTIKWMDKLCINNTTHHCIDECNRNCLCFDRVYKQKEEEMNSIK 647
DB 591 HNFELWVYTYLLRDTIKWMDKLCINNTTHHCIDECNRNCLCFDRVYKQKEEEMNSIK 650
QY 648 KLEFTKKNIOGSYSNNINLEGEYFFKWMKLDKDEAKKWELENIKKRKNFESULENN 707
DB 651 KHEK-----YDGSITGN---NDYFVLENVLD-----ELRQDTEAAGNSKIGIK 697
QY 708 DYL-----ENAILLDHLKETATICKDNNTNEACETSH-NATNPKYKPR 752
DB 698 DTLAKKRTQAADATEQKNTIDLFYEDSEAEKCK--KIOBECQPKPTKVRNPKY--- 752
QY 753 GGIQPTNINIKELQYFRSAVEEA-RNRGLHLKLGKRAHEGILYKRGKGRKQFKNLCIMI 811
DB 753 GNNTYDALAGVAAKLOQEAKEQDLRDSNSALKAANSQGSYNSQGPDDFKKMLGKITQ 812
QY 812 KHSNRNLGFSNGPCDGCYKCDGIRYTHKSDNDRVNNEDYKPPMGVKNPTNITVLS 871
DB 813 KHSNA-IGDSKNPCNNNG-----KERNVGEKMNNGE-YKMSITDILTPRRRHFTCSN 865
QY 872 LEHLQTDHPLNGNIYDVLVNSFLGVLLSAKYEANKIIRMYEKNKNLGKPEVTPDKH 931
DB 866 LEHLNTR-----STGLTSDKAHSLIGDVLIAAK-----KEGBDIK--TLTENDN 909
QY 932 OTTCRAIRYSFADIGIILINGRDLWENGMVAKIQGLELYVFGIHSIKSGKGDYKND- 990

Db 910 RSSICRTMKYSPADIGIINGTDLMDJNGDATGVONNLKIDFISKITEELKROHPDKEND 969
QY 991 ----DABKYLKRENNMWEANRAKAYWEAMKCDIKYLDKSGHSTOSSGYSHTPLDXY 1046
Db 970 DKYTNDKHTKRLSDMWEANRDQYWKMTCPTR-----NGNIQCG---ATPDDY 1016
QY 1047 IPQKLRMTWAEYCCVQVQKKEDEKKECKECKDKNDGCGCTRESGCTGCTEACNEY 1106
Db 1017 IPQKLRMTWAEYCCVQVQKKEDEKKECKECKDKNDGCGCTRESGCTGCTEACNEY 1106
QY 1107 NDIJLMEQMNIIISDKYKELHEQAOQMSVNSGIEASTKKNHIDRVIETSELYQNG 1166
Db 1070 KTIQOPADQMNELSNKYQLYIQAIAALNGTEKSTTTKDKDKNVIDEQLKHEAN 1129
QY 1167 GKSNGSGTDES-----AVIGTNTYENGVAYLHDGTFDCCOSONEFCDEKSDGK 1218
Db 1130 G-----TKRPPPEAHDRPRRAATSKSDVYETTAGYIHOEARTEGJGQVFCNNNG---- 1182
QY 1219 NEKAFRDKPOHDGACGCGSSKPTRVQIKTKKAEKDEKTEKYVNDILKENDGKKOYE 1278
Db 1183 NMEYAFSLPHEKYMCKC-----NENKASS-----PEELGSDSF 1218
QY 1279 DCHPRKNSNGYDPMOGINLVEDPRVCMPPRROKLCVHFILANDNEIKKLQSGVNLKEAF 1338
Db 1219 DDH-----QTPREDEBH----- 1232
QY 1339 IKSAAETFPSPWYKSKDGEKNELDKELKEGKIPALISMFTYTGDRDLFTGDISK 1398
Db 1233 -----SSEGEDESEDEKE----- 1248
QY 1399 GHGEGSKLEQIDSLFKNDGOKSPNGKTRQEWMTESHEIWEAMLCALYIGAKKDDPTE 1458
Db 1249 -----EVEBHVDGADKA----- 1262
QY 1459 NNGYNNVKSSTSTLEEFKAPQFLRWLTWYDYCYTRKYLKDVQEKCKSNDQKCD 1518
Db 1263 ----- 1262
QY 1519 TBCNKKCEDYKMKKKKKEMIPQDKYKDERDKKRRDROHIGWMTDYGTNATDYLNK 1578
Db 1263 ----- 1262
QY 1579 FTASGDKPBGASVVOIRIOLLEKQAYYDADHCGCTFIENDKYTNISSDKCKGLVK 1638
Db 1263 ----- 1262
QY 1639 EANTGAIKQKQPNYNNYLKELTEDVLEPSSRLRICFHALDGYNTDPEVKDENGILRRKL 1698
Db 1263 -----GAVSQPEASPT-----TKDVV----- 1278
QY 1699 MEVATEGYNIGQYKKEKKEKIKITSAHKSYEVPPCSAMKSYFYDLROIIGIDNLE 1758
Db 1279 -----KPPCDIYK-----ELFSSNVDYLO 1296
QY 1759 DEKQTEENLKIFNKNGTSVGKSGDSTTGNPGSTARKEFWNENKECVNANAMIGCYKGR 1818
Db 1297 ----- 1296
QY 1819 DDGNGSNAKSPEDLKCCGSVPDDDYPMGKNRDEGTAYQFLRWFAMGEDEFCHEKEL 1878
Db 1297 -----KACST-----KYGPKN----- 1308
QY 1879 EKLVCANDYTCGDNEDRRKCTOACQYKFFISEMKPOYEQKIKKYGENDKIYSEHPV 1938
Db 1309 ----- 1310
QY 1939 AKDADAREYLDKOLKIKENKSGDCEYKCMKDVSTQRLTDGNSONMPASLDEPKEVEG 1998
Db 1311 -----WRCIPKTSNDVTEBDQ----- 1328
QY 1999 KQNCQVPRPRVRETPSPKYSLSKATASKKEAKTAPPKQPKYENLTTEMAQOTRT 2058
Db 1329 -----GSRHVASTPE-----SGSNSDKNGATCIPRRRRRLYYGKL----- 1364

QY 2059 RRAAOOTRRKRTSATTTESDVGIMVAKAISNKPDSNGIGECNPKTYGOYPRWGCIWKS 2118
Db 1365 -----EOMANKHNTETVSOGAEAT-----EARG-----SEAPAPG-----G 1395
QY 2119 KENENIGCMPPRRKRLCINNIOYLANETENKRDNDIKAEFICAMIEFOFLAKYIE-- 2176
Db 1396 KRESSG-----GKETPSDK-----LRTAFISAAVEFFELMDRKTEWL 1434
QY 2177 -----NPAENELONGTIPDEKFRIMTYGYGYKDMFEFTDI 2213
Db 1435 AOKKAELONGIDLYSSGDGDPDN--ONKLLNGVILPPEFLRMLFYLDYDILVHGN 1492
QY 2214 SMDKLIITYNSVTIILNENNNKKKODKKDELRKI-----FEK 2253
Db 1493 TDSGNTGNSNNNNIYLEASGKEDMOKIOETIOELPRNGSTPLVPKSSAQTPDKWNE 1552
QY 2254 NKKFTWEGMIGLYY-----HLDENEKKEKIDNY----- 2283
Db 1553 HAEIWKGMICALITYTEKNPDTSGARDENKIEKDEVEYKFFGSGADKHGASTPTGYK 1612
QY 2284 -OYN-DMTKL-----TPSLEPYKRPQFLRMTWAEYFCNKRKQLE 2327
Db 1613 TOYDEYKYLEDTSGAKTPSASIDPPLSDFYLRPRYRYLEWQNGNCKERRKRLQIK 1672
QY 2328 AGC-----KEYECNG-----SNDGKTQCAEACVYQYOMFIKMKTEYEROREK 2370
Db 1673 EECMGSGDKKTYGDDGEOCDRDTSNESVADLEBGRCSGSCFRYKWKIRKREKEDQANA 1732
QY 2371 FKDKDGGKYYDYPSTERDIEKATCAHEYLMLKLELGNKDCSGOMKPSOLPRTTQOS 2430
Db 1733 YSKOR--TKYE-----EGS 1744
QY 2431 QSSDANDMPESIDYVPEEFNKECEPELSKSGSMITHKITEPKIPANCVEKAAAY----- 2485
Db 1745 KGAGLND-----HNEFCVKLG-----TCYDAAPAFNLRLKN 1775
QY 2486 -LSKEAEN-NMDITL--KKEFPIESTKESKSWTNNPCDPKPPAPADKYIGRN 2539
Db 1776 GPCKKDNENGNDINFGTEETFEFRAENCK-----PCSSFR-----INCRN 1816
QY 2540 -PCENRENNRKYVDEMYKCYNSKIFYOEKKRVCVPRRREHMLNLDKIERLKSNTL 2598
Db 1817 GNCNSGDG-----DTKEKC----- 1830
QY 2599 LKAVRTARNEGIDILIKFNSENGCAAMPICDTMKYSFADLGIVRGDMIRIGIYLPV 2658
Db 1831 -----NGGITTGNTGNTGCT-----EDVV----- 1851
QY 2659 EIKLYVFEYIYKWRNKNKGRNRYNDVQTFERSAMWDANRDKAMTCKAPEDAKLERK 2718
Db 1852 -----MHVSDKANAEFEGLDEACENAGIFGIRDEK--CRKYGSLHICQ 1698
QY 2719 GRMDGF--ERITLIQDKCGHKDPPVDYIPOFRMWTMSEYCCALMELEKFKKSC 2775
Db 1899 EKGNGALNDQIILVR-----ALIKRWVEYF-----LEDYKIKIKKL 1935
QY 2776 DHCKTSDRCKNDYDENKCEQCTRCOEYKNFVLYKSLFDIOSKRYKELY- 2825
Db 1936 KPC-----JENGSGTICNGKKNKRVGWMYIKLKDWTIKINNHLEKKNKBDKN 1986
QY 2826 -----EOPITYKISTYD--HYQNFVOKLKT--FKSECVSEFSEYIAHETSKLWYFN 2874
Db 1987 VTSLVTVNLETVLOIAANDKREQTSLDKLTSLGNCNCPENS-----R 2030
QY 2875 ENOSSNIRIYAFETPRTSKYKACSCYLPBSKNPLDNCPTQONKDGCKELOTFFPGSKNDY 2934
Db 2031 KNDGENE-----DALDCML-----NKLETKIHECKTOH 2058
QY 2935 DNNDLNNNAVILNNSDD--NKGVLTPRRRHLCYTPITAYNYRRGDKRILKLLTISA 2991
Db 2059 ENSVEN-----SHQPHRNGCGNPP----- 2078

```

QY 2992 FSQGLGQYKSEELCEPAMVSYADYSDIIKGTMDMTSLSEKIKLFTSNATEEN 3051
DB 2079 -----DEEDLLEE----- 2087
QY 3052 RKTWMENNRQIWHAMLCGKIKATSKVTLDGEGCOLPKDETNOPLRWLEMAKQACKER 3111
DB 2088 -----BNPVEQ-----PGCPPTQOPEPD----- 2107
QY 3112 KHVSLSLTKCPSPNEDNEFASELLNPOGCONDIRKYISLILIKNTMENTINIKYKOLKO 3171
DB 2108 -----DKCGLEKKEKKE-----Q----- 2124
QY 3172 OSSGNIDNKPSEENVOSYIKSDSCALDELINELVYTGKNNENNEKEVILKLYGLY 3231
DB 2125 -----EQPAEDGGAIVPSP-----PG-- 2142
QY 3232 FVEDETHKNHVLGNKKEEQGTVPKALYFTFPHVDSFYQAPLFSTHRVAYDPKNDILK 3291
DB 2143 -----SEPEADKGVKPAET-----PKP-----QEPDLSHPA-----VIP 2173
QY 3292 SSISVIVSALGL-----IALHFKKKRKS SV-DLRLINIPQEGYGMPTLESKNRYIYR 3346
DB 2174 SLVSTSLAMS VGI GFAFYFYLYLKTKSSVGNLFQIOLPKSDYDIPKLSNRYIYPT 2233
QY 3347 SGYKRGKTYIYMGDT-----SGDEKYMMDLSSDITSSSEYBELINDIYVPGSPRYKT 3403
DB 2234 SGYKRGKRYIYLEGDSTGYTDHY-----SDITSSSEYBELINDIYVPGSPRYKT 2287
QY 3404 LIEVLEPSKRDIPSD-----DTPS-----NDPRTNRFIDDEKNEL 3440
DB 2288 LIEVLEPSGNNTTASGNNTTASGNNTTASGNTPSDTQNDIONDIPSKITDNENOL 2347
QY 3441 KHDVSOYL-----PTEPNNNYKSADIPMTEPTLYSDNPEKPFITSIHDLTYGKEI 3497
DB 2348 KDEFISQYLOSEPNTPE-----NMLGYNDNTHPTSH-HAVEEKPFIKSHIDRLNLFSEXY 2404
QY 3498 SY-----NINNSTN-----TNDIPMAANDSYGIDLINDSLVYVLLIY--M 3539
DB 2405 NYMFGNSGNPNINISDTSNMDSLTSNHSPYNDKLDYSGIDLINDALSGHIDYDEM 2464
QY 3540 MK 3541
DB 2465 LK 2466

```

RESULT 14
P90580 PRELIMINARY; PRT: 2647 AA.

P90580: 01-MAY-1997 (TREMBLREL. 03, Created)
01-MAY-1997 (TREMBLREL. 03, Last sequence update)
01-DEC-2001 (TREMBLREL. 19, Last annotation update)
DE FCR3-VART11-1 protein (Fragment).
GN FCR3-VART11-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FCR3;
RX MEDLINE=97154495; PubMed=9001213;
RA Hernandez-Rivas R., Mattei D., Sterkers Y., Peterson D.S.,
Wellens T.E., Scherf A.;
RT "Expressed var genes are found in Plasmodium falciparum subtelomeric
regions.";
RL MOL. Cell. Biol. 17:604-611(1997).
DR EMBL: U67959; AAC47438.1;
DR InterPro: IPR004258; Pfam:
Pfam: PF03011; Pfam: 2.
FT NON_TER 2647 2647
SQ SEQUENCE 2647 AA: 303263 MW: 65FD700611348BC7 CRC64;

Query Match 13.8%; Score 2666.5; DB 5; Length 2647;

Best Local Similarity 23.6%; Pred. No. 1,6e-121;
Matches 869; Conservative 379; Mismatches 861; Indels 1635; Gaps 120;

```

QY 31 SRSARNVLERYAKNIRHPEKAKEHVS-----LKGLDTKEPFGG-----PSRPVKN 79
DB 9 TNSAKAEVLDEIGETIO-----KKAHSDADTFPSOLKNGEKKFTNGEITMOPNSKL-- 61
QY 80 HNYYPYPCNLDKHEHTNLRYDVNLRHPCGHEQNRNFEDESEEC-GNKRIRYKRRKND 138
DB 62 -----CELDHTIDTNDVGHSN--PCGQGTVPFRPDNNSQCTKNKIKD-SYDNSV 109
QY 139 IACAPRRRHHCKNLEBALDIN---TONIHLLGVNLTAKEGESYVNNHP--HGT 192
DB 110 GACAPYRRLHLCSHNLSEIOTNNYDSKASHNLAEVCYAAKEGESIYKNEYQLGHTT 169
QY 193 SDACTLLARSFADIGDIVRGIDMEKRPVHD-----KVEYGLREVPKIH----- 236
DB 170 EGICTALARSFADIGDIVRGIDMEKRPVHD-----KVEYGLREVPKIH----- 236
QY 237 DGMEDVYKNDYNDPGSGNYYKLEBAMNNVNRNKAWEATCDASYSKGYPMQSESNTPLES 296
DB 230 NGAEEERYK-----DGSNYYKLEBAMNNVNRNKAWEATCDASYSKGYPMQSESNTPLES 284
QY 297 NPKCGHKGQVPTNNDYVPOYLWPMDEGEEFCKRIRIKAKTKDCR--NDKERLYCS 353
DB 285 GERKQCIDGTVPNNDYVPOYLWPMDEGEEFCKRIRIKAKTKDCR--NDKERLYCS 344
QY 354 HNGHDCCTTYIMKGGILHLDKCTDCSTKCYFVFWJLNOQEARFKOKEREKIQSYLSN 413
DB 345 RNYGDCCTTYIMKGGILHLDKCTDCSTKCYFVFWJLNOQEARFKOKEREKIQSYLSN 404
QY 414 DNKFVNNINSEYKQYKLETOYATNDPFLNLNEGKYCKGGLPGEKIDTFTNSADK 473
DB 405 SONSPPKNNMYETDEFGNLKDKQYQSWD--FLKLINSETPCTNIDAKSKIDFTK--DPE 461
QY 474 GIFYRSYQVCPDQCY--CDGIXTKHSDNDRERANNNDYRPMQVKTNTLYVLSGN 531
DB 462 ETSFHEYCDPCWCGCTQAD--TWKRLYENDPQCEIKRYRPPGVETEDVLYTGK 520
QY 532 EOGDITOKLENFENSSYNYADKNNQKWEYCKDENIKRCLTEONTLEINDNPKIISFHN 591
DB 521 ENKDIYKLEPCKTGNTEFK--NEENWCYQYGN--DKVLENGEELGGER-KYKQVDNF 577
QY 592 FELWTVYLLDRTIKWMDKLTCTCJINNTTTCIDECNRNCLCFRHWVAKOKEEMNSIKLFT 651
DB 578 LMFVYVAMLDSDIEMSKNCLSKDKCTCITTCNDQCQYDRIKWKYHMTQIKHFD 637
QY 652 KKKNIQ--QSYYSNNINLFEQ--YEFKVDKLDKDEAKKELEMINIKRKNFESNLENNRDY 709
DB 638 KQIDPQGWGHFYVLETVLESDQFFDTTKAYGDRRELVHIOEMQKKBEV--LHEDASN 695
QY 710 LENAIELDLHLKETATTC-----KDNV-----TNEACETSHNATNPCKVPRGTOPTK 759
DB 696 MKTIIDLDLHELKAKQCLVNHKDNCPADLSQSEDEEDIDQROKCAKP--SGTHIRA 754
QY 760 NIKETIAYFKRSAYEERANGL-HKLGKAHEGITYKGGRRKDKMLCIMIKHSRNL 818
DB 755 LVNRYVSNMHEKRRKRLDYNVGSKRLGDAKAEKRSCTTIKLD--ICSIITDHSNAKR 813
QY 819 GFNSPGDGGTGGDGIOTRVRVGVDEPHEMKHDEYVIMPRRHHICTSNLEHLOT 878
DB 814 GHTDQCKRDKSVNNVNR-----RMDDTAGFTISNTKYDLYMPKRRHFTCSNLEYLOT 868
QY 879 DHPLENGNIVD---DLVNSSELDVLSAKYEANKIIRIYKKNLKLCKPKEVTPKHOTJI 935
DB 869 NKLINGNDINGNPNINISDTSLGVLFAANEADEIRIKMYKNQ-----DYKDNATI 919
QY 936 CKAIRYSFADIGIIRK-----DLMERGDAVYKLGHELYVGNITHSLGK 982
DB 920 CRAMKYSFADIGIIRK-----DLMERGDAVYKLGHELYVGNITHSLGK 971
QY 983 KGNDKYNDADPKYLLKENNWEANRAKVMAMKCDIKYLLDKDSGHOSSTOSYCYSDHTP 1042

```

Db 972 -----YKEDIP-YTKLEDWEANRKKIWEAMOC-----PPNGSPFCKSYHIG 1014
QY 1043 LDDYIPQRLKMTWEMAMCYKQVQKEXYDLKECKECKDKNDNGGCKJESGTGOTKTEA 1102
Db 1015 LDDYIPQRLKMTWEMAMCYKQVQKEXYDLKECKECKDKNDNGGCKJESGTGOTKTEA 1069
QY 1103 CNEANDIIGL---WKEOWNIISDKYKELHQAQSVNSNGSIEASTAKNHIIDRNVLEFLS 1159
Db 1070 CKHYKILKNLILHKEQWDMKEIKYLLYLOAQTTANANGPDYISGLVDENEKPYVAFLE 1129
QY 1160 ELYOANGK-----SNKSGTSDSAVIGINTYENYGAALH-DTGPFDDCOSONEP 1209
Db 1130 ELYKENGKIGNRDPTRAKRSKRETAAPASVANDYSTAGVYHOMEGPMHCKCTOTEP 1189
QY 1210 CDEKSDGKNEKXAFRDKPODHGACGCKSGSKPTRYQITKR-----KAEKO 1258
Db 1190 C-EKTBOYENENTFKNPQYKDACICNTRPPPKEDSRRSDSDEEVEKTEYKEEKA 1248
QY 1259 TE-----CKTWNILKENDGKQVDECHPKKNSNGP-DWOC--- 1294
Db 1249 TEDAVDTGPPPAKEATTYLDVCPVAGVLT---KENLEMACPTKYGPRAPTSMKCIPT 1304
QY 1295 -----GNINLYEDPR-----VCMPPRROKICV---HELANON 1323
Db 1305 EKTNAATGSGSGGNGALQRAKATVYESGSPVTSNGSICIPPRKRLYIOKLHMAAGN 1364
QY 1324 EIKKLOSOV-----NKEAFIKSAAEFFSWYXXYSKSGDEGNEIDKEL 1367
Db 1365 TVVSGAOPPOGCTSSPSGKETPPSDKLTAFIOSAAIETFFFLMDRYK---KGKATAKE 1420
QY 1368 KEKG-----IPAPFLSMWYTTGDRFLPGTD 1395
Db 1421 KKKQWDSYPLSTADPHNPPVSLVAPNPNYKTCVIPPFLRQMYTTLGDYADLFFGKN 1480
QY 1396 ---ISKHGE-----GSKLBQIDSLFRNGDOKSPNGRTQEWMTESHEIWEAMCAL 1446
Db 1481 DLYIDTRKNGDKDAERKKIKDAIERVLKNADSOPPSDKRLTWBQNEBHIMNGICL 1540
QY 1447 VIGAKKDOFTENYGNVNNKFSDKSTLLEEFKAPROFLMLETWYDDCYTROKYLKDVQ 1506
Db 1541 T----- 1541
QY 1507 EKCKSNDOLKCDTECKNKCEDYVYKMKKKKEMIPQDKYKDERDKRPHROHIGVWYDY 1566
Db 1542 -----YK----- 1543
QY 1567 TGTNATDYLNRKFTYASCGDKPGSASVQORNIQLLEKQAYYADKHCCTKFTENDKYYTN 1626
Db 1544 -----EKDEKGT- 1550
QY 1627 ISSKDKCKGLVKEANTGAIKMQNKPNNYNNLKELTEDYLPFSRRLICFHALDGNITDP 1686
Db 1551 ----- 1550
QY 1687 EYKDENGILKRLMEVAATEGYNLGOYKKEKKEKIKTSDAHKYSYEVPPCSAMKSYFD 1746
Db 1551 PLKQEGIKSALMD-----EKNKKPKQD---KKQY----- 1577
QY 1747 LRDILIGDNLDEKOKTEENLKKIFNNKGS---VKGSDSTTGNPGSTARKFFMNNEN 1802
Db 1578 -----DKVKIDEN---SGTSPKIVPAPRPPTTTPPSPSTSF----- 1612
QY 1803 KECVWNNMICYKRGDDGNGSARSDEDLKKGGSVSDDDYMGKNRDRGTAYOFLRW 1862
Db 1613 -----SRP-----PFYR 1620
QY 1863 FAENGEDECKHKEELEKVLGACNDYTCGDNEDRKKCT---DACTQYKFFISEMKPOYE 1919
Db 1621 LEWAEFFCERBRKRLERI---KVEOMBEDGKKOKSGSGEGCEERK----- 1665
QY 1920 KQIKKYGKMKKIITSEHPVAKDADAREYLDKQKLTICENKSGDCEYKCKMDVSTORLTD 1979
Db 1666 -----ODYSTVRDPYCPBECGKYCR----- 1684

QY 1980 GNSQNNPASLIDEPKEVECKCNCQVPRGPRVRRRETPSPVSLISKATASKKTAAPT 2039
Db 1685 ----- 1684
QY 2040 KQPKVENLTTEBRAQTRTRRAAQOTRKRTSTATTESDVGTWYKALLSNKPDSSNGJEG 2099
Db 1685 ----- 1684
QY 2100 CNEPTYGOYPKMGCIYVSKSENGENICMPRRKKILINNIOYLANETENKRDNIKEAFI 2159
Db 1685 -----FKRW---IGKKKDEYD---KQK---AYNOKTARANN----- 1715
QY 2160 KCAALFTQPLMKYIILEENPAENELONGTIPDEFKRIMYYTYGDYKDMFPGTDSNDKI 2219
Db 1716 -----NDNAF 1720
QY 2220 ITVNSVTTLN---ENKKKODKKDE-ELRKIFWEKNNKFFIWEGMIGLYHFLDE 2273
Db 1721 STTLDTCTAGDFLTGLNKGPCKNNDNDSGENKIF-----DE 1759
QY 2274 NKEKIRDNQYNDMTKLTPLSLEEFVKRPQFLRMETWAEFCNKRKEQLKLAGCKEY 2333
Db 1760 N-----GDTFKYT-----QYCG-----TCSLNGF 1778
QY 2334 BCNGSNDGTQOCACACTYONFIKKWTEYERQREKPKKDKKDYDSTERDIEKA 2393
Db 1779 KONGD-----CVRTRN----- 1791
QY 2394 TCAHEYLNNKLKELCGNKDCSCMOKPSSQLPKTTQOSQSDANDMPESLDVYEPFKCE 2453
Db 1792 TC-----NSGNRTTITADI----- 1807
QY 2454 CPELSKKSMTHTKITEPKIIPACVEKAYYLSKEANN---DTLKEKFIPIESTK 2509
Db 1808 ---KNG-----GNSAENMLVSDDJNSGNGFNDLBA-- 1835
QY 2510 EKESKSWTNNPCDPKPYAPDKYIGRNPCCNRENRFP--VDYEMCKYKSKPFQEK 2567
Db 1836 -----CKN--ANIRGIKENKWC---YFCK 1857
QY 2568 KRCVPPRREHMLRNLDEIKIERLKDNYLLKAVRTARNEGIDILKNSENGCAMNP 2627
Db 1858 SDVC----- 1861
QY 2628 ICDIMKYSFADLGDIVAGTDLRIGYLPPEYELIYVFEYIYGKNNKGRKNTYDVQ 2687
Db 1862 -----GLKKNNDI- 1869
QY 2688 TERSAMWDANKDITWAMTKAPEDAKLFRKGRMDGERITLLIODKCGHKDDPPVDYIP 2747
Db 1870 -----DONOILIRAL-----FKR----- 1883
QY 2748 QRFWMTEMESEYCKALMELEFKFKSCDCKTSDCKNDYDENKCEQCTRCOEYKRFV 2807
Db 1884 -----WLEYF---LDYINIRIKKLNPC-----INNGEKAICINGVE--- 1917
QY 2808 LKMSLEFDIOSNKYKELYEOPITYTKISTYDVONFYOXKLTFFSECSVSFSYLMETS 2867
Db 1918 QW-----INHKRTETNLSF----- 1933
QY 2868 CLWYKENENDGSSNIRYAFETPKSYKACSGCTLPSKNPLDNCPTDOKNGCKEIQTF 2927
Db 1934 ---NEOYNGDPTERNPLRFEVDLIRQ---IATIDK-----GHNHGVLKLYKSV 1977
QY 2928 FCSKNYDNNLDWNNAYLVANSDDKGVLLIPRRRHLCRPTITAYVNRKGEILKKKL 2987
Db 1978 KC-----NCGN-----NSQNGKEGENDVLCI 2000
QY 2988 LTSAFSGOLLGQYKSEELCEAMKYSYADYSIILKGTMDNDSLSKIKKIIFTSNE 3047
Db 2001 L-----QKLEKKAKKC-----KDNPEYS-----GIDPOOPE 2026

```

QY 3048 ATENRRTWNNRNRQIMHMLGCKYKATATSKVTLDEGMCQIPKDEETNOFLRWLEMAKOA 3107
DB 2027 VSENNH-----IDEE---EQPLEEEN-----T 2045
QY 3108 CKEKKVSDSLKTCPCRSNEDNEASELLRQPCQNDIRKYLINLIKMTMENLIKRYK 3167
DB 2046 VEHKRICDVLKKNHNRQNER-----LVKRPVLOPRLTKR 2081
QY 3168 QLKQSSGNIDNKPSEBNVOSYIKSKDSOCALBLNDINEITGTKNNEKPEVKLYK 3227
DB 2082 KKKKKRKKKKK-----NODFH---PRHL 2103
QY 3228 PGLFVEDETHKHNVLGDNKE---EEQTVPRKALYFTPHVDSFYQAPLESTHRVAQDP 3285
DB 2104 PCGAFINTNPKTKTPSSGKNPWEHPAVLPA----- 2135
QY 3286 KNDLKSSISVIVSALGLALHFKKKFSSVDLRLINLPOGEYGMPLTESKNRIRY 3345
DB 2136 ---LVTSTLAMSIGIGPAFTYFLAKKTKSTIDLILSL-IPKSDVDIPKLSPNRIRY 2191
QY 3346 RSGPYGKTYIYEGDT---SGDEKYMMDLSSDIT-SESEYEELINDIYVPGSPKY 3401
DB 2192 TSGYKRGKRYILEGSGTSGYTDHY-----SDLTSSSESEYEMDINDIYVPGSPKY 2245
QY 3402 KTLIEVVLPS-----KNDIPSDDTPSNDTPRTNFIIDEMWELK 3441
DB 2246 KTLIEVVLPSGKLSGNTIPTSGNNTTASPTQNDIPIDSPRPIT-----DDEWELK 2298
QY 3442 HDVVSQYLPTPEEN--NNKYSADIPMTEBPTLYSD-----NPEKPFYISI 3486
DB 2299 HDISNNLQNPQDVPDRDYSGSSSTNTTTSRQVNDVNDNTHPRKRNHVDQKPFITSI 2358
QY 3487 HDRDLTYGKEISTYNKSTNTNDIPMANRNDSTRCIDILINDSL 3530
DB 2359 HDRNLYTGEYNYNVNM-VMTMDIDIPINSHNVYSGIDILINDTL 2401

RESULT 15
ID 094657 PRELIMINARY; PRT; 2212 AA.
AC 094657;
DT 01-FEB-1997 (TEMBLrel. 02, Created)
DT 01-FEB-1997 (TEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Erythrocyte membrane protein 1.
GN PFEML1.
OS Plasmodium falciparum.
NC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5833;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=JTG.
RA Yang S.O., Wollish W.S., Gut J., Wu J., Ahn J., Petersen C.,
RA Fujioaka H., Akawa M., Isech J.H., Nelson R.G.;
RA "The molecular cloning and DNA sequence analysis of Plasmodium
RT falciparum erythrocyte membrane protein 1."
RL EMBL: U31083; AAB06961.1;
DR EMBL; J01083; AAB06961.1;
DR InterPro; IPR004258; PFEML1.
DR Pfam; PF03011; PFEML1.2.
SQ SEQUENCE 2212 AA; 253594 MW; A33456648C852FC CXC64;

Query Match 13.8%; Score 2677; DB 5; Length 2212;
Best local Similarity 31.1%; Pred. No. 3.7e-121;
Matches 713; Conservative 307; Mismatches 800; Indels 476; Gaps 83;

```

```

QY 139 IACAPRRRRMDCKNLBALDIN---TONIHLLGNVLTAKYEGESIVNNHP---HKGT 192
DB 110 GACAPYRRHLHCSHNLESISTQNNYDSSKAKHNLAECYAAKEGESIVANNYEOIGHHTT 169
QY 193 SDACTALARSFADIGDVRIDMKRPVND---KVEFLREVEKTIH----- 236
DB 170 EGICITLARSFADIGDVRIDMKRPVND---KVEFLREVEKTIH----- 229
QY 237 DGMEDVEYKNDYNDPDSGNYKYLREAMNNVNRKNKWEAITDASYSKYGFQMSNPPLFS 296
DB 230 NGAEBERYK-----DGSNGYKLRDWMNNANRLDIWKMIKICAPONAQYFRNTCSNGEKP 284
QY 297 NPKCGHOGKVPYPLDYPVOTLWEDWGEFCKRRIKAKYKDSR---NDKERLYCS 353
DB 285 GERKQCIDGVPPNIDVPPYOLWFEEMAEFCKRRIKAKYKDSR---NDKERLYCS 344
QY 354 HNGHDECTTITMKGILHLDNCKTDCSKCYEEFWLGNQOEAFAKOKEKYETOSYLSN 413
DB 345 RKGIDCTKTRISDKYMANBCTKCLYCDPYKAMIDNKKKEFKOKKCBENETRYNNES 404
QY 414 DNKFVNNINSEYKQFYEKLETOYATNDFNLNLEGYCKGGLPEKDIITFTNSADK 473
DB 405 SQNSPKYNNMYETEDFIGNLAKVOQSMND-FLKLNSETCTNIDKRRIDFTK--DPE 461
QY 474 GIFYRSEYCOVCPDQGYK--CDGIKTYHKSNDREVRNNEDEYKPPWGVKPTNITLYSGN 531
DB 462 EYESHTEYCDPCWCGKGLKQADG-TWRKLYENDPCQIFKPEYEPKEEPEIDVLTGK 520
QY 532 EGGDITQKLENFCSSTNYNDKNNQKWEYKDEINRKNLEONTETINNNPKIISHNF 591
DB 521 ENKDIYKLEFCKTQDNTGK-NEENNYCYOEN-DKVALNCEELGGEK-KKQDNE 577
QY 592 FELWVLYLSDYTKMNDKLTQJINNTTHCIDECNRCLCFDRVAVKQKEBMSIKKLYT 651
DB 578 IMFVVAHMLKDSIEMBSKLSNCKLSDKKTCKKNDCKCKEYKIGKKYEMTQIKKHPD 637
QY 652 KKKNIQ-QSYYSNINNFEG-YFFKVDKLDKDAKKELEMKKRNKNEFSNENRDY 709
DB 638 KQTFQOQMGREYFLEYLEGDQFPTDTTKAYGDAARELVHIOEMQKKREOV--LHEASN 695
QY 710 LENAIELLDHLKETFATIC---KDNNTNEACETSHN-----ATPVCVPRGCTOPTK 759
DB 696 MKTIDELDLHELKKAQCIYVNHKNDNCPDLSSEDEEDIPORONPCAKPSSABRAL 755
QY 760 NIKELIAYFKRSAYEABRNRL-HLKGAHEGYTKRGRKRDKNLCIRIMKSNRNL 818
DB 756 -VNVASNMHKKKRLQVNVGSKLKGDAAKGYRRSGTTRKLD-ICSTINDHNAKR 813
QY 819 GFSNGPCDGGKGTGDTOTRFVGTIEWEYDEPNKRDIEDVIMPRRHICTSNLEHLOT 878
DB 814 GHTDQPCRK---DSKSEMFTEBDGMR-PAGFISKTYKDIYMPRRHCFYSNLEYLOT 869
QY 879 DHPNGNIYD---DLYVNSFLGDVLSAKYEANKIIMYKREKNLKLKPKETDCKHOTI 935
DB 870 NKLNGNDINGNPNFINDSEFLGDVLFANYTEADRIKMTKKQN-----DKDNATI 920
QY 936 CRAIRYSFADIGDILIRGRDLMERNGDVWKLQGHIEYFNGHKLKSGNDKYNDAPKY 995
DB 921 CRAMKYFADLGDILIKTDMDNDNGSGSKTRDKARELFDITKKKHPI-KEIYEDRP-Y 978
QY 996 LKLENNWEANRAKAWEMAKCDIKYLLKDSGHSTQSSYCGYSDHPPLDITPOKLRMT 1055
DB 979 TKLREDEWENRKRKIWEAMOC-----PPNCSFPCSKYHTPLDIDYIPORLRMT 1027
QY 1056 EMAYEYKQVQKKEQDKKECKECKDKDNGGCKESGCTGCTECNECNEINDIIGLKE 1115
DB 1028 EMAYEYKQVQKKEQDKKECKECKDKDNGGCKESGCTGCTECNECNEINDIIGLKE 1084
QY 1116 QWNISIDKYELHQAQMSVNSGIEASSTAKNIDNVLEFLSELYQONGK----- 1168
DB 1085 QWMEIKYKYLVAQATTAANGSPDIYSGYLDENKRPVNFLEFLKRENGGKIGNRDT 1144

```



```

QY 1169 --SNKSGTSDSAVIGINTYENAGAYLH-DTGNFDCQSONEFCDEKSDGKDNKAYFR 1225
Db 1145 PRAKRSKRETPASAYAKNDVYSTAGYVHOEJGPHMECKTQTEFC-EKTDQYNNENYFK 1203
QY 1226 DKPODHGACGCKSGSKTRQIKTKK-----KAEKDE----- 1260
Db 1204 NPPQYKDACICJONRPPPKEDSRKSEDSDEBEKVKETKEKATEDAVDTGPPAPAE 1263
QY 1261 -----CTVNDILKENDGKROVEDCHPKKNSNGYP-DMOC-----CNI 1297
Db 1264 TTTLIDVCGIVAGVLT-----KENLENACTKYGPAPTSKCIPTTEKTAATGSESSGNG 1319
QY 1298 NLVEDPR-----VCMPRRQKLCV--HFLANDNEIKRLQSQV----- 1332
Db 1320 ALQRAKRAFVSGSPVTSNGSICIPRRRLRYIQLHDMASGNTVSGOQTPGCTSS 1379
QY 1333 -----NLKEAFKSAAEFPFSWYKSKDSEGNELDEKKE----- 1369
Db 1380 PSGKETPSDKLRTAFIOAALETFFLMRYKKE-----KEIEKKEKVVANGGLVPSLNGP 1435
QY 1370 -----GKIPAFILRSMTYFGDYRDLFTD---ISKHGE----- 1402
Db 1436 PQOPGVGDSPOSKIQCGVILPPEFLRQMFYTLGDIADIFGKNDIYDITKNGDKDIAR 1495
QY 1403 GSKLEQJDSLFKNGDQKSPNGKTRQEWTEHSEIWEAMLCALY-----KI 1449
Db 1496 EKKIKDAIERVYKKNADSQPPSPDEKQOTWMBQNGEHINMGICALYKKEKDEKGPILKONE 1555
QY 1450 GAK-----KDPETEYGYNNVYFSDKS-----TLEEFAPKRPQFLRWLJE 1489
Db 1556 GLKSLMDEKMKKPPDOYQYIDKYLKLDENSGTSPTKNDHPPTPLNTISRPPYRTLEE 1615
QY 1490 WYDDYCYTROXYLADYDEKCKSND--OLKCD-----TECNKRC 1525
Db 1616 WGEFECRERKRLEKIKYECMEDEGKQKCSGDGECEIRKODYSTVRDFYCPGCKYK 1675
QY 1526 EDYKYMKKK-EMIPQKYYKDEKDKRFRQIHGVAVTYDTGTNADYLNKFTASC- 1583
Db 1676 RPYKRWIEKKDEYDKQEAANNQCTDARNNNDNAFSTLTCTAGDFLOTLNGPCK 1735
QY 1584 ----GDKPGSASVQORNTQLEKQAYYADAKHCGCTKFIENDK---TYNISKDKCKGL 1636
Db 1736 NDWDDSGENKKIPDENGDTRKYIYCGT---CSLNGFKCNGDDCRVRYNTV---CMGS 1788
QY 1637 VKEANTGAIKWONGPNYNNMLKELTEVDLPSSRLRICFHALDGNVYTDPEYKENDGLRK 1696
Db 1789 NRTTITFADDIKNG--NSAEINMLVSD-----DINSNGFN- 1823
QY 1697 RLMEVAATEGINLGQYKEKEKEKIKTSDAHKISYEVPPCSAMKYSFYDLRDTIL---- 1752
Db 1824 ---DLEACKNANIFKGIKENWK-----CYVFCSDVCGLKKNNNDIDONOILIRAL 1872
QY 1753 ---GIDNLEDEKQTEENMLKIFNKGTSVSGSDSTGNPSTARKFFWNNKCEVWNA 1809
Db 1873 FKRLLEYFLDYNKIRKTLNFCINN-----GKAICTNGCYBQ-----WINOKTEWT- 1920
QY 1810 MICYKRRGRDNGSNGSARSDEDLKKGSVPSDDYPNGKRNDEGTAYOFLMFAEWGED 1869
Db 1921 -----NIKN--RFNEQY-----NGDTEM-KSFSRSLVLDLIRQIA---AT 1955
QY 1870 FCKHKEKELELYGA--CNDTTCGDNEDKRRKCTACTQYKKFISEMKPQYKQIKKYE 1927
Db 1956 IDKGNHNGLVKLVSVKN---CGNNSONGKE-----GE 1986
QY 1928 NKDKIYSEHPAKDAEDAREYLDKOLKICEN--KSGDEYKCMKVSTQRLTD----- 1979
Db 1987 ENDLVLC-----LLQLEKAEKCKDNPERISGIPQPC--EVSBNHTEDEQPLE 2034
QY 1980 --GNSQNNPASILD---EPREVECKCNOYRGPBPVRRRETSPRVSILSKATASKKEA 2033
Db 2035 EEBENTVEHPKICDDVLTQPOPEERGETCEBSPGTDVKEEHEEK-----KEEKDEEB 2089
QY 2034 KTAPTKQPKVENILT 2049

```

Db 2090 PGLPPTPPAPASST 2105
 Search completed: April 28, 2003, 10:35:24
 Job time : 233.729 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 35.7348 Seconds

(without alignments)
2916.372 Million cell updates/sec

Title: US-10-087-013-2

Perfect score: 19407
Sequence: 1 MGFSCKYFIIMGNASSLE.....IDLINDSLVNLILIMKY 3542

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

al number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3489	18.0	3060	2	US-08-487-826B-14
2	2972.5	15.3	2710	2	US-08-568-459A-12
3	2972.5	15.3	2710	2	US-08-487-826B-12
4	2972.5	15.3	2710	4	US-09-210-288-12
5	1743	9.0	2182	2	US-08-487-826B-16
6	1032.5	5.3	700	2	US-08-568-459A-10
7	1032.5	5.3	700	2	US-08-487-826B-10
8	1032.5	5.3	700	4	US-09-210-288-10
9	633	3.3	921	2	US-08-568-459A-8
10	633	3.3	921	2	US-08-487-826B-8
11	633	3.3	921	4	US-09-210-288-8
12	617	3.2	1435	2	US-08-568-459A-4
13	617	3.2	1435	2	US-08-487-826B-4
14	617	3.2	1435	4	US-09-210-288-4
15	431	2.2	749	2	US-08-568-459A-6
16	431	2.2	749	2	US-08-487-826B-6
17	431	2.2	749	4	US-09-210-288-6
18	396	2.0	10182	4	US-09-134-001C-3159
19	381	2.0	3135	4	US-08-323-170B-2
20	381	2.0	3135	4	US-08-954-441-2
21	365	1.9	411	2	US-08-568-459A-20
22	365	1.9	411	2	US-08-487-826B-32
23	365	1.9	411	4	US-09-210-288-20
24	361	1.9	3248	1	US-08-353-700-1
25	361	1.9	3248	5	PCT-US95-16216-1
26	343.5	1.8	1115	2	US-08-568-459A-2
27	343.5	1.8	1115	2	US-08-487-826B-2

28	343.5	1.8	1115	4	US-09-210-288-2	Sequence 2, Appl
29	343.5	1.8	1115	6	5198347-6	Patent No. 5198347
30	340.5	1.8	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
31	291.5	1.5	2482	1	US-08-328-254-6	Sequence 6, Appl
32	279	1.4	362	2	US-08-568-459A-18	Sequence 18, Appl
33	279	1.4	362	2	US-08-487-826B-30	Sequence 30, Appl
34	279	1.4	362	4	US-09-210-288-18	Sequence 18, Appl
35	276.5	1.4	2285	4	US-09-308-375-2	Sequence 2, Appl
36	276	1.4	411	2	US-08-568-459A-19	Sequence 19, Appl
37	276	1.4	411	2	US-08-487-826B-31	Sequence 31, Appl
38	276	1.4	411	4	US-09-210-288-19	Sequence 19, Appl
39	273.5	1.4	311	2	US-08-568-459A-21	Sequence 21, Appl
40	273.5	1.4	311	2	US-08-487-826B-33	Sequence 33, Appl
41	273.5	1.4	311	4	US-09-210-288-21	Sequence 21, Appl
42	257	1.3	3418	3	US-08-755-587-44	Sequence 44, Appl
43	256.5	1.3	1663	5	PCT-US93-07261-16	Sequence 16, Appl
44	251	1.3	3418	2	US-08-639-501-2	Sequence 2, Appl
45	251	1.3	3418	3	US-09-044-946-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-487-826B-14
Sequence 14, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chittis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: A35
ATTORNEY/AGENT INFORMATION:
NAME: Israelisen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 3060 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-487-826B-14
Query Match 18.0%; Score 3489; DB 2; Length 3060;
Best Local Similarity 26.7%; Pred. No. 1.5e-240;
Matches 1054; Conservative 472; Mismatches 934; Indels 1490; Gaps 148;

QY 9 IIMGNASSLEBDAPKSPIIKESHKSARVLEYAKNIRHSPK-YAKEHVDSIKGLDTRKA 67
 Db 4 LAKMGPEAEAGDDIED-----ESAKHMFDRIGKOVYKVEEKEKREKQIGRLSEA 56
 QY 68 EFRGSPSTPVNKNINYYPPYPCNDHKENTLRDQVNLHPCGREGREONFDEDESEEC-G 126
 Db 57 KFEKNESDPOTPEB-----PCDLIDKHYNV-----TJNVIPFCADRSVPRSDYGGQCTH 108
 QY 127 NKIRYKRRKNDATACAPRRRRHCDKNLEALNDINTONIHDLGNLVYTKGESIYVNN 186
 Db 109 NRIKDSQOGNKGACAPRYRLHYCDQNLBQIEPIKITHNLHNLVYDCAKAFEGQSTIOD 168
 QY 187 HP-HKGT-----SDACTAARSAJDIGDYRGIDMEKPNVHD-----KYETGLREYFKKI 235
 Db 169 YPKYQATYGDSPSOICTMALARSPADIGDYRGHDLVGNPQEIQRQOJLENNKTIFFGKI 228
 QY 236 HD-----MEDEVKNDYNDPDSGANYKTLREAMWNNRKNKWEALITCDASYSGYMOSESMT 292
 Db 229 YELKNGAEARYND--PE-----FKLREDWMTANRRTWKALTCNA-WGNTYF-HATYCNR 280
 QY 293 PLESPNKGCHKOGKVPJNLIDYVPOYLRFWDEWGEFRCRRNIIKIKYKDCSR-----NDKE 348
 Db 281 GERTKGYCNCNDQDPYTYEDYVQYLRWFEEMAEDPCRRKKNKIKYKNCCKKEDD 340
 QY 349 RLYCSINGHDCITTYIKKGLHLDNCKCTOCSTCKYFEVWLNQOBAFKKQEKYEKIO 408
 Db 341 R-YCSHNGYDCEKTKRAIKRLRYGKOCISCLYACNPNYMINQKQOFKOKKYYDEIK 399
 QY 409 SY-----LSNDNRKPVNNINSE-YKOFYEKLEKETOYATNDPFLNINSGKYC-----454
 Db 400 KXENGASGSGSRKRDAGGTTTNYDGYEKKFYDELANKSEYRYVDKLEKLSMEICTRYK 459
 QY 455 --KGLPGEKIDYFTNSA-----DDKGIYRSEYQVCPDQCGKDIKTYH-KSND 504
 Db 460 DEBGGITDKVNNSTSGASGTNWSOGTFYRSKYCQPCPYGVK-----KVNNGSSME 515
 QY 505 RERAVN-----EDYKPPGVKPTNTYVLSGNEGDIQOKLENCN-----SS 547
 Db 516 WBEKNNKGKSKLYEPKPKDEKGTITTLTKSGKHDDIEKLNKCKDEKNGDTINSGGG 575
 QY 548 TWYKDKNN-----OKWECYKDENINCKLEONTEINDNPK-----II-----586
 Db 576 TGGSGGSGNSROELYEEMWC-YKGEDVYVGHDEDEDEYEVKNNAGCILKKNKKE 634
 QY 587 -----SEHFEELWYTYLLRDIKANDKIKTI-ANTTTHC-IDBCNRNC 629
 Db 635 ECGNTSEKPEDEIOKTFNPFYVWAMHMKDSIHKKKLOKLOLQNGNRLKCGNKNCCND 694
 QY 630 LCFDRVYKOEHEWNSIKKLTFTKKNIO-----QSYYSNINLFEGYF-----672
 Db 695 BCFKMTIOKKKEMWKIYOHF-KTONIKRGSGDNTAELIPPDHYVLOYNLOEFLKGD 753
 QY 673 -----FVYMDKLDDEAKWELMENIKRKNKFNLENND-----YLENAIELL 718
 Db 754 SEDASEESENSIDAEEA-----ELKHILREIIESEDNNOASVGGVTEQKNIMDML 807
 QY 719 DMLKFTATTC-----KDNNTNACELSHNATINPCYKPGGQO-PTKNIKETIAO 766
 Db 808 NYEKEDADICLIEHDEEERKKGNGCTEGENFRYVNPSCGESGNRYPLANKVAYO 867
 QY 767 YFKRSAYEARNRGLHKRAHEGIRYKGRKRDKDNLCRIMKHSNMLGSGSCD 826
 Db 868 MHHKAKQLABRAGSALRGDISLAQFKNGRNGSLKGOICKINENYSDRSGSGGCT 927
 QY 827 KCGKGDGLOTFRVVTEWEVDEHMRKDEHYIMPRRRHICTSNLEHLOTDHPLNGNI 886
 Db 928 GK-DGDHGVMRIGTESNIEGKKQTSYKNVFLPPREHCTSLNLDV-----GSV 980
 QY 887 V-DDVYVNSPFGDVLISKYFANKTIIRMYKKNMLKGYETDP-----KHQTTICRAIYS 942
 Db 981 TKNDKASISLGDVQLAKKTDAAELIKRYKDONNI-----QULTDPIQOQDQAMCRAVYS 1036
 QY 943 FADIGDILRGDLWRNGDNVKLQGHLETYVGNIHKSLKG-KGNDKYND-----APKYLKL 998

Db 1037 FADIGDILRGDLWRNGDNVKLQGHLETYVGNIHKSLKG-KGNDKYND-----APKYLKL 998
 QY 999 RENWMEANRAKYWEAMKCDIKYIKDKSGHOSYSGYSDHTPLDDYIPQKLRMTEMA 1058
 Db 1097 RADWMEANRHQYWRAMK-----ATKGIIC-----PQMPYDDYIPQKLRMTEMA 1141
 QY 1059 EYTCYQKKEKYEKLEKCECKDKNQGCCTYESGTGCTKEACNEYNIDILGKEQWN 1118
 Db 1142 EYTCYQKKEKYEKLEKCECKDKNQGCCTYESGTGCTKEACNEYNIDILGKEQWN 1199
 QY 1119 IISDKYKEHEQAOQSVNSNGSIEASTAKNHID-----RNVIELTSELQOONGK-----SNK 1171
 Db 1200 KISDKYNLLYLAKTSTNPG-----RTVLGDDDPYQKQVDTLTIHKAASIAARYLARA 1255
 QY 1172 SGTSDSEAVIGNTITYENVGATLHDTGNFDCQSONEFCDEK-----SDGKDNKVAFR 1225
 Db 1256 ASPEIEMAAADITYSTAGYIHOEIGGQOQOTCEKKNHATSTSTKENNEYTFK 1315
 QY 1226 DKPOHDGACGSGSKPTRVQIKTKKAKEKDTECTYVNDILKENDKQOVEDCHPKN 1285
 Db 1316 QPPEIATACDINSQTE-----EPKKEENYESACKIYEKLEKNGRTVYGCNPKRES 1371
 QY 1286 SNGYEDMOC-GNINLVEDPVCMPPRROKLYHFLANNEIKKLOSOYNLEAFIKSAA 1344
 Db 1372 ---PYDMOCKNNDISHD-GACMPPRROKLYHFLANNEIKKLOSOYNLEAFIKSAA 1427
 QY 1345 EFFEYVYTKK-DEGHELDKELKEGKIIPALFSMFTYFGDYRDLFGTDISKGHBG 1403
 Db 1428 EFTLSMOYKSKNDESEALILR-----GLIPSOFLSMATYFGDYRDLFLNDISKQNDV 1483
 QY 1404 SKLEOISLFPKNGDQSPNGKTROEMWTEHSHETWEMALCALYKAKKD-----FTENY 1460
 Db 1484 AKAKIKIKFISKDSKSPSGLSROEMWTKGPELWKMALCALYKVTDTNKKRKNKY 1513
 QY 1461 GYNNYKESDK-STLIEFAKRPQFLRWLTWYDYCYTRQYKLDVQDECKS-NDOLKCD 1518
 Db 1544 SYDKVNOSONNPISLEFAKRPQFLRWLTWYDYCYTRQYKLDVQDECKS-NDOLKCD 1603
 QY 1519 T-----BCNKCEDYVYK-KKEEMIPQDKYKDERDKRPRROHIGVWDTYGTNATDY 1574
 Db 1604 DAKHCKNACAYOYEVANKKFEFGOTNNVLANVOPOPEKGYEYK-----1654
 QY 1575 LNRFTASCGDKPGSASVYVQNIQLEKQAVYDADKHCQKTFENDKYTNISSKDKC 1634
 Db 1655 -----GVOP-----IGNEVYLQK-----CDNNK-CSC-----1676
 QY 1635 GLVYEANTGALKMKNKPNYNNIKELTEDVLFPSRLRLICFHALDGNYTDPEYKENG 1694
 Db 1677 -----MDGNVLSVSPREK-----1689
 QY 1695 RKRLMEVATGEYMLGQYKKEKKEKIKTSDAKHYSYEVPPCSAMKYSYDLIDLIGI 1754
 Db 1690 -----PPGKT-----1699
 QY 1755 DNLEDEKQTEENLKKITFNKNGTSVKGSDSTGNPGSTARKFPMWNEKCVNNAICGY 1814
 Db 1700 -----1699
 QY 1815 KRGDDNGSANSASDDLLKCGSVPSDDIYPMKKNDEBTAYOFLWFAWGEDFCCHK 1874
 Db 1700 -----1699
 QY 1875 EKELEKLVGACNDYTCGDNEKRRKCTDCTOYKXKFISEMKPOYERQIKKYGKNDKIYS 1934
 Db 1700 -----EKC-----1702
 QY 1935 EHPVADAEAREYLDKOLKICENKSGDEYKCMQDVSTORLTDSNQNMPASLDEPK 1994
 Db 1703 -----DC-----1704
 QY 1995 EYEGKNCQYPRGPRVRRRETPSPRYSLSKATASKKEAKTAPPTOPKRVENULTTEMA 2054

APPLICATION NUMBER: US/08/568,459A
 FILING DATE: 07-DEC-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Israelson, Ned
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NIH121.001CPI
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-8550
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2710 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium falciparum
 08-568-459A-12

Query Match

15.3%; Score 2972.5; DB 2; Length 2710;
 Best Local Similarity 26.5%; Pred. No. 1.3e-203;

Matches 865; Conservative 442; Mismatches 873; Indels 1083; Gaps 120;

QY 9 IIMGNASSLEEDAKSPITIKESHKSRANVLEKAKINIRHPSK-YAKEHVSILKGLUTRA 67
 Db 6 LAMGPPEAKGGDIED-----ESAKHMFDRIGDYVDYKVEEKEKGGKQIGRLSEA 58
 QY 68 EFRGSPSTPVNKNHYYPYPCNLDHKEHTLRVDVNLNHPCHGRQNPDEDEESEC-G 126
 Db 59 KFEKNSDDPOTPRD-----PCDLHKYHTNV---TTNVINPCADRSVDVSESDYGGQCTH 110
 QY 127 NKTIRNKKKADAIACAPRRRHKCDNLBALDINTQNIHDLGNLYVAKYEGESIVNN 186
 Db 111 NRIKDSQGGNKGACAPYRLHWCQDNLQIEPIKITHTNHLVDCMAKREGOSITDQ 170
 QY 187 HP-HKGT-----SDACTALASPADIGDIVRGIDMKRPVND-----KYETGLREVEFKI 235
 Db 171 YPKQATYTGDSPOQITMLARSPADIGDIVRGIDLYGNPOETKORQOLENNLIKTFGKI 230
 QY 236 HD---GMEDEVKNDYNDGSGNYTKLREAMVNNRNKWEATITCDASYSGYFMOSESNT 292
 Db 231 YKTLNGAEARFYND--PE---FFKLREDMWTANRETVKALITCA--WONTTF-HATCNR 282
 QY 293 PLPSNKGCGHKGKVPNTNDYVQYLRWPDENGEEPCRRNITKLKVVDSR-----NDKE 348
 Db 283 GERTKGYCCANDQVFTYDYVQYLRWPEWAEDEFCRRKKNKIDVYKNCRGKDEKDD 342
 QY 349 RLYCSHNGDCTTTIMKGLILHDKCTDCSTKCYFEVWLGNOQDAFPKOKYEKEKIO 408
 Db 343 R-YCSHNGYDCERTKRAIGKLYRGKQCSCLACNPYDMINNOEQDQKAKKIDDEIK 401
 QY 409 SY-----LSNDKRYNNINSE-YIKQYETKLETOYATNDTFLNLNEGKYC--- 454
 Db 402 KYENAGSGGSRKACGTTTNTYDGYEKKEFYDELINKSEYTVDFKLKLSNEICTKVK 461
 QY 455 --KGLIPEKIDITFTNSA-----DDKGIPIRSRYCGOVCDGCKGCIKTYTH-KSND 504
 Db 462 DEEGTIDFKVNNVSDTSGASSTVNESQITFRSKTCQPCPCGVK---KVNNGSSNE 517
 QY 505 RERVNN-----EDYKPPWGVKPTNITVLYSGNEQDITQKLENNFC-----SS 547
 Db 518 WEEKNNGCKSGKLYEPKPRDEGTTITLKSQKGDIDLEKLNKFCDEKANDITNSGSG 577
 QY 548 TNYKAKNN-----QKRECYKKNENIRKCLEONTEINNDNR-----IT----- 586
 Db 578 TGGSGGSGNGOEYEEKCC-YKGEDVYKVGHDDEDEYENKAGGLICLTKNQKKKE 636
 QY 587 -----SPHNFELWVYLLRDTIKMNKTKCI-NTTTHC-IDECNRNC 629
 Db 637 EGNNTSEKEPEIQTIPFTFYTYVAHMLKDSIHMKKRLQCLQNGNRIRKGNKNCND 696

QY 630 LCFDRVYKQKEEWSNISKILFTKKNIQ-----QSYSNININFEYF--- 672
 Db 697 ECFKRIITQKDEWKTIVOHF-KTONIKRGSGSDNTAELLIPDHVVLQYNQEEFLKCD 755
 QY 673 -----FKYVNDKLDKDEAKWKMELMENIKRKNFSLNNRD-----YLENAIELL 718
 Db 756 SEDASEEKSNSIDAEBA-----ELKHLREITSEEDNNOBASVGGVTEQKNIMDKLL 809
 QY 719 DILKETATIC-----KDNNTNEACETSHNATPYCVKPPGGTQ-PPKNIKELAQ 766
 Db 810 NYEKDADICLETIEDEEEKKQGNNECIEGENFRVPCGSEGGKRRPYLAKVAYQ 869
 QY 767 YFKRSAYEERARRRGLKLGKAHEGIIYKRGKRPKDNLCRIMKHSNRNLGFSNGPCD 826
 Db 870 MHHKATQULASRAGSALGDISLAOFKNRNGSTLKGQICKINENYNDNSRNGSGPCT 929
 QY 827 GKGTDGIOTRFVYGVTEWEVDEPHNRKDHEDVIMPPRRRHICTSLLEHLOTDDHPLNGNI 886
 Db 930 GK-DGDHGGVRRIRIGTEWMSNIGSKQTSKANYLPFRREHMCSTNLNLDV-----GSV 982
 QY 887 V-DLVNNSFLDQVILSAKYEANKIIRMYKEKNILKGRVYDP---KHQTTICHAIRYS 942
 Db 983 TKNDKASHSLDQVLAQKTDAAEIIKRYKDQNTI---QLTDPIDQKQDQAMCRARAYS 1038
 QY 943 FADIGDIINGRDLTFRNDGMVYKLGHELVFGNTHKSLKG-KGNDKYND---APKYLK 998
 Db 1039 FADLDIIRGRDMDDEKSSIDMETRLITVRKNIKEDHGILKDPKRYTGDSKKPAVKYL 1098
 QY 999 RENNWEANRAKWEAMKCDIKYLDKDSGHSTOSSYCSYSHPTLDDYIPQKLRMTEMA 1058
 Db 1099 RADWWEANRHQVYWRMC-----ATKGLIC--PGMVVDYITIPRLMRTEMA 1143
 QY 1059 EMYCVQKEKYDKLKEKKECKDNDGCGCTKESGTCTCEACNEYNDIILGMKQWN 1118
 Db 1144 EYCAQAOEYDKLKLKICADMSKGDGK-CP-QGDVDCGKKAACDKYKEIEKEMNQR 1201
 QY 1119 IISDKYELHQAQMSVNSGIEASTAKNIID---RNVIIEFSELVQOQNGK-----SNK 1171
 Db 1202 KISDKYNLYLQAKTTSTNPG---RYLGDGDDPYQOAMDFLPIHKASIAARVLKRA 1257
 QY 1172 SGTDSQVAVTNTTYENGVAYLHDTGNFDCQSONEFCDEK-----SDGKNEKYAFR 1225
 Db 1258 AGSPLEIAAARPIPYSAAGYIHQETIGYGGCQOTQFCECKHIGATSTSTKKEKETF 1317
 QY 1226 DKPDHOCAGCGSKGSPTRQVITKKAAREKDECTVNDIILEKNGKQVDECHPKKN 1285
 Db 1318 QPPEVYATACDCINRSQTE-----EPKKKEVSEACKIVERTILEGKNGRTVEECNPKES 1373
 QY 1286 SNGTPDMQC-GNINLVDPVRCMPRRQKLCVHFLANDNEIKRLQSOVNLKEAFISAA 1344
 Db 1374 ---YBDMDCNNINIDISH-GACMPRRQKLCVYIHAESEGTENIKNDNKLAKDFITAA 1429
 QY 1345 ETFESWYIYYSK-DGEGNELDKELKEGKIPPAFLRSFYTFGYRDLFGTDSIKSHGSG 1403
 Db 1430 ETFISWYIYSKNDSEKILIR---GLPSQFLRSWYTFGYRDLICNTLDSIKRQNDV 1485
 QY 1404 SKLEQIDSLFKNGDQSPNGKTRQEWMTESHEIWEAMCALVKGAKKD---FTENY 1460
 Db 1486 AKADKIGKFFFSKDGSKSPGSLROEWMKTNGBEIKGHCALTKLYTTIDNKRKTKNDY 1545
 QY 1461 GYNNYKFSK-STLLEFAKRPQFLMWLTWVDYCYTRQKLYKDOEKCS-NDQJLKD 1518
 Db 1546 SYDKVNDQSONGNSLEFAKRPQFLMWLTWVDYCYTRQKLYKDOEKCS-NDQJLKD 1605
 QY 1519 T---EKNKCEDVYKYM-KKKEMIPQDKYKDERKAKKFRQNHGVMTDYTGTVNADY 1574
 Db 1606 DAHNRQACRAQOYEYENKKEFSQGTNNFVLKAVQDQDEYKQYKXD----- 1656
 QY 1575 LNRKFTASGDKRGSASVQNRNIOLEKQAYYDADHCGCTFIENDDKYTNISSKDKC 1634
 Db 1657 -----GVQD-----IQGNEYLLQK---CDNNK-CSC----- 1678

```

QY 1635 GLVKEANTGAIKMNKCPNNYNNLKEITEDVLPSPRRRLICFHALDQNTDTPVEKDENL 1694
Db 1679 -----MGNVLSVSPKEK----- 1691
QY 1695 RKRLMEVAATEGYNLGQYKEKKEKIKITSDAHKYSLEYVPCPSAMKYSYDLRDLIGI 1754
Db 1692 -----PRGKY-----AKHY----- 1701
QY 1755 DNEDEKOTEENLKIKIFNNKSTSVGKSDSTTGNGPSTARKEFFWENKECVNNAICGY 1814
Db 1702 ----- 1701
QY 1815 KRGRDQNSGNSARSDDELKCKGSPDDYIPMGKNRDEBTAQFLWFAKMGEDPCKKH 1874
Db 1702 ----- 1701
QY 1875 EKELEKLVGACNDYTGCDNDRKRCCTACTQYKKFISEWKPOYKQIKRYGENKDIYS 1934
Db 1702 -----EKC----- 1704
QY 1935 EHPVADAEDAREYLDKOLKICENNSGDEYKCMQDVSTQRLTQNSQNMPSLDEPK 1994
Db 1705 -----DC----- 1706
QY 1995 EVBGKNCQYPRGPRVRRTPSPRVLISKATASKKEATAPPTQPKVENLTTEMA 2054
Db 1707 -YQK--HVPSP-- 1725
QY 2055 QTRTRAAQOTRRRTSTATTESDVTGTMVRAILSNKPSRSGIEGCPNKTYGYQ--PKWG 2112
Db 1726 -----PEAPVTVYDVCISIVTLK--DTNNFSDACGLK-YGKTAPSSWK 1766
QY 2113 CT-----VGSKENENGICMPRRKLCINNIOYLANET-----ENKRD 2151
Db 1767 CIPSDTSGAGATGKSGSDSGSICIPRRRLRYVGLQ--EMWATLPQEGAPSHSRA 1824
QY 2152 NDKKAFIKCALETQFLMKYIIE-NPAEN-----LONGT 2188
Db 1825 DDLENAFQSALETEFLMRYEERKPOGDSQOALSQLTSTYSDEDEPPKLLQNG 1884
QY 2189 IPDEFKRIIMYTYGQYKDMFFGDISNDKRIITVNSVTTLNENNRKODKKDEELRK 2248
Db 1885 IPPDFLRLMYYTLIGDYDLIVHGNTSDSGNTNSNNNNIYLSGSKEDMOKQOEIEQ 1944
QY 2249 I-----FEKNKFTWEGMTYGLTY-----HLDENEKEKIR 2280
Db 1945 ILPRNGSTPLVPKSSAQTPDKMNEHESIMKMICALITYTEKNPDTSARGDEKRIEKD 2004
QY 2281 DNY-----OYN-DMTKL-----TPSLEEFYKRP 2302
Db 2005 EYEKEFGSTADKHGTASTPTGYTKTYDYEKYKLEDTSGAKTPPSASSDPPLLSDFVLRP 2064
QY 2303 QFLRMFTMAEEFCNKRREOLKLEAGCKEYE-----CN---GSN 2339
Db 2065 PYRYILEMONGNCKKKHHLAQKHCKYEBENGSGSRGIGITROYSGDEACENMLPKN 2124
QY 2340 DGTQOE-----CAEACVLYONFIKKMWTXYEROREKFRKXD----- 2376
Db 2125 DGYVPLEKESCAKPCSSYKRWIESKEFEKQEKAYEQKDKVNGSNKHDNGFCETLT 2184
QY 2377 -GKKYKXY-----PSTRDLEKATCAHEIYLMKLELCGKKDCSCMKPSSQLPKTQOS 2430
Db 2185 TSSAKQFLTLGPCCKRNNVEGKITFDQDTFK-----HTKDDPCLKFSVCKR----- 2234
QY 2431 QSSDAMPSPSLDYVPEEFKCEPELSKSGSMITHKTEKIPKIMNCVEKAAAYLSKEA 2490
Db 2235 -----DECD-----NSKGTDCRKN-----NSID-----ATDI 2266
QY 2491 ENNDITLKEKFLPIESTKESKSNSTNNPCDPKPKVADKYIGRRNCPEN----- 2543
Db 2257 ENGVDSTV-----LEKRVASDSKSGFNGD-----GLENACRGAGIEGI 2295
QY 2544 REBRFVVDYEMKCYKNSKRYOEKKRVCVPRRREHMLRLNDELKIRLSDNSYLLKMR 2603

```

```

Db 2296 RKD-----EMKC-KRVGCV-----VKRP----- 2313
QY 2604 RTARNEGIDILKNNSENGCAMPICDTMYKSPADLADIVRGTDMLRIGGLPVEIKLY 2663
Db 2314 -----ENVNG-----AKGHIIDIRALVR-----RWV 2336
QY 2664 KVEYIYKWRNKNKNGKRYND-----VQFRSAMWANDRDIKAMTCKAPEDAKLERKG 2719
Db 2337 EYFFEDYKIKHSHIRKNGEISPCKIKNCYKWDQKRRE-WKEITEREKD----- 2387
QY 2720 RMGFEHITLQDKCGKDDPPDYIPQFRNMTESSEYCCALMELKFKKSCDHCK 2779
Db 2388 -----QYKNDSDDDNRSFLETLLIPQTTDN-----ANKKYIKLSFGNSG-CS 2432
QY 2780 TSDRCKNDYDENCCEQCTKQCEYKKNFLKMSLFDIOSKRYELYEQPIYTKISTYDHY 2839
Db 2433 AS---ANQONKN-----GEKDAI-----DMLKRLND-----KIGCEBK 2465
QY 2840 QNFVQIKLTKFSKCSVSSESEYLIHETSKLNYKFNENDSSNIRTYAFEBTPKSYKE-- 2896
Db 2466 HH-----QTSDTCESDTPQPTLED--ETLDDIETEERKKNMFKICENVLTKAQODE 2518
QY 2897 -ACSTLPSKNPLDNCPTDQNKD 2918
Db 2519 GGC---VPAENSEPAAATDSGE 2538

RESULT 3
US-08-487-8268-12
Sequence 12, Application US/084878268
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsep, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO

```



```

Db 1702 -----EKC----- 1704
QY 1935 EHVADAEADAREYLDKOLKICBNKSGCEYCMKDVSTQRLTDGNSQNPASLDEPK 1994
Db 1705 -----DC----- 1706
QY 1995 EYEGKCCOYRGPBRVRRPSPRVLISKATASKEATAPPTKQPKVENTITEMRA 2054
Db 1707 -YOGK--HVPISIP-----PPPVQPO----- 1725
QY 2055 QTRRAAOOTRRKRTSTATTESDVGTVMKAILSNKPSRSGLEGCPNPKTYGY--PRWG 2112
Db 1726 -----PEAPVTVDVCSIVTLTK--DTNPSAGCLK-YGKTAPESSWK 1766
QY 2113 CT-----VGSKENENGICMPRRKKLCINNIOYLNET-----ENKRD 2151
Db 1767 CIPSDPKSGAGATWGKSGSDGSGICIPRRRRRLVYGKIQ--EWATALPQEGAPASHSRA 1824
QY 2152 NDIKEAFIKCAIETQGLMKYIIE-NPAANE-----LQNGT 2188
Db 1825 DDLRNAFIQSALETFFLMQRYEERKPGQSGSOALSQLTSTYSDEEDPPDKLLQNGK 1884
QY 2189 IPDEFKRIIMTYTGYDYKDMFPGTDSNDKLIITVNSVTTLNENNNKODKKDEELRK 2248
Db 1885 IPPDFLIMFTYTGDIYHGNISDSNGTNGSNNNNIVLEASGKEDMKQKQEKIEQ 1944
QY 2249 I-----FWEKNNKTIWEGMIGLVY-----HUTDENEKIR 2280
Db 1945 ILPRNGSTPLVPKSAQTPDKMWHESIMKMICALTYTEKNPOTSARGDERKIEKD 2004
QY 2281 DNY-----QYN-DMTKL-----PPSLEEFKRP 2302
Db 2005 EYVEKFPFGSTADHGTASTPTGYTKOYDEKVALLEDTSACKTPSASSDPLPLSDVFLRP 2064
QY 2303 QPLRMFTMAEFCNKRREOLIKLEACKEYE-----CN---GSN 2339
Db 2065 PYRFLYLEMGONCKKKHKLQIKHECKYVENGGSSRRGGITROYSGDEGACEMLPKN 2124
QY 2340 DGTQOE-----CABACTYTNFIKKMTEYEROREKFKKDKD----- 2376
Db 2125 DGTVPDLKESPCAKPCSSYKRWIESKGEFEKQKAEQKQKCVNGSNKHDNGFCETLT 2184
QY 2377 -GKKYDY-----PSTRDLEKATCAHEYLMMKLKELGCKKDCSCMKPSSQLPKTQOS 2430
Db 2185 TSSKADDELTLGPKCPNNVEGKTIFFDDKTFK-----HKKDDPCLKFSVNCCK----- 2234
QY 2431 QSSDANDMPESLDYVPEEFKCEPELSKSGSMITHKKTIEPKIPMNCVKAAYLSKEA 2490
Db 2235 -----DECD-----NSKGTCCRNK-----NSID-----ATDI 2256
QY 2491 ENNMADTLKKEKPIPIESTKEKESKNSWTNNPCDPPKPYADKYIGRRNFCEN----- 2543
Db 2257 ENGVDSTV-----LEKRVASDSKSGFNGD-----GLENNACRGAGIEGI 2295
QY 2544 REENRFKVDYEMKCYKSKRYOEKKRYCVPRRHEHMLRLNDEIKIRLDSNVLKMR 2603
Db 2296 RKD-----EMKC-RNVGCVY-----VCKP----- 2313
QY 2604 RTARNEGIDILIKNFSENGCANPICDTMKYSFADLGIYRGTDMLRIGLYLPEVEIKLY 2663
Db 2314 -----ENVNGE-----AKGRIIILRLVY-----RWV 2336
QY 2664 KVEEYIYKRRKNNKGNKKNYND-----VQTRSAMWANDKRDYKAMTCKAPKEDAKLRKG 2719
Db 2337 EYEFEDNKIKHSHIKNGEISPCIKNCVEKVKVDOKRKE--WKEITERKD----- 2387
QY 2720 RMDGFERITLIDKCGKDDPVDYIPIQFRWMTWSEYCYKALMELEKFKSCDHCK 2779
Db 2388 -----QYKNSDSDDNVRSFLETILIPQITDAN-----AKNKVTKLSFGSGC-CS 2432
QY 2780 TSDRCKNDYENKCEQCKTRCOEYKNFLVKWSLFDIOSNKKYELBQPIYTKISTYDHY 2839
Db 2433 AS--ANEQKN-----GEYKDAI-----DCMLKKLKD-----KIGCEBKK 2465

```

```

QY 2840 QNEVQKLTFRKSCVSFSEYLTHTSKCLNYKFNENDGNSNIFTVAEETPKSYKE--- 2896
Db 2466 HH-----QTSPTCECSDPPQDPTLED--ETLDDIETEAKKNMPPKICEVNLKTAQOODE 2518
QY 2897 -ACSTLPSKNPLDNCPTDONKD 2918
Db 2519 GGC--VPAENSEPAAATDSGKE 2538

RESULT 4
US-09-210-288-12
; Sequence 12, Application US/09210288
; Patent No. 6392026
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xia-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA: Patent Release #1.0, Version #1.25
; APPLICATION NUMBER: US/09/210, 288
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH121.1EMDVI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-09-210-288-12

Query Match 15.3%; Score 2972.5; DB 4; Length 2710;
Best Local Similarity 26.5%; Pred. No. 1.3e-203;
Matches 865; Conservative 442; Mismatches 873; Indels 1083; Gaps 120;

QY 9 IIKGNASSLEGDASKPIIKESHKSARNVLEARYAKNIRHPSK-YAKEHVDLSKGLDTKA 67
Db 6 LAKGPREAAGGDIED-----ESAKHMDRIGKIDYQYKVEAKERGGAGLGRLSEA 58
QY 68 EFRGSPSTPVNKHNYTTPYPCNDLKHETNLRKYDVNLRHPCRGHREONRDEDESESC-G 126
Db 59 KFEKNESDPQTPED-----PCDDLHKYHTVY--TTNVINPCADRSVRFSEDEYGGCCTH 110
QY 127 NKINRYKNDALICAPRRRHMCCKNLDELNDINTQNIHDLNLYTAKYEGESIYNN 186

```

Db 111 NRIDSOOGDNKACAPRYRLHVCDQNLQJEPKITNTNHLVYVCMAMKFEQSITQD 170
 QY 187 HP-HKGT-----SDACTALARSFADIGDIYVIGIMFKPNVHD-----KVENGLEBVEPKKI 235
 Db 171 YPKYQATYGDSPQICITMLANSFADIGDIYVGRRLYLGNDPEIQROOLENNLKTITGKI 230
 QY 236 HD-----GMEDEKKNYNDPGSGNYTKLRLAMVNNRNKWEATITDASYKSGYFQMSBNT 292
 Db 231 YEKINGAEARYGND--PE---FPKLEDMWTANRETYWKATITCN--WGMTYF--HANCNR 282
 QY 293 PLFSPNPKCGHKGKGYPTNLDPVQYILMPEDWEGEGRKNKIKLAKKDSR-----NDKE 348
 Db 283 GERKKGCRCDNDQVPTYPYVPOYLKMEBMAEDPCKRNKIKDYKRNCRGDKDKD 342
 QY 349 RLYCSHNGDCTTITMKGLIHLDNKCTDCTCKVEVWLGNOQAEAFKKRKEKEKEIQ 408
 Db 343 R-YCSRNGYDEKTKRAIGRLYKQICISCLYACNPYVDMINNOKEFDQKKYDEIK 401
 QY 409 SY-----LSMDNKFVNINSE--YKQPYELKETOYATNDTFLNLNKGKVC--- 454
 Db 402 KYENGASGSRQRKRDAGGTTTTYNDGEKKEFYELNKESEYRYVADKFEKLSNEEICTKVK 461
 QY 455 --KGLPEKEDITFTNSA-----DDKGIYRSEYQVCPDGVKQDGIKTH--KSDND 504
 Db 462 DEEGGIDTFKNNVMSDTSIGASGTVNESQGFYFSKTCQPCPYGVK---KVNGSSNE 517
 QY 505 RERYNN-----EDYKPPMGVKNPTITVLYSGNODGDTOKLENFCN-----SS 547
 Db 518 WEKRNKCKCKSGKLYEKPDKEGTTITILKSGKHDIIEKLNKFCDEKNGDITNSGSG 577
 QY 548 TYNKDKNN-----OKMECYKDEENINRCKLEONTEINNDNR-----IT----- 586
 Db 578 TGSSGGNSRQELYEKMC--YKGEDVYKVGHDDEDEDEENKNAAGLICILNOKRKE 636
 QY 587 -----SFHNEFLMVTYLLDITKNDKIKTCI--NNTTTHC--IDECNRNC 629
 Db 637 EGGNTSEKEDDETOKTFNPFYVVAHMLKDSIHMWKILQRCLOQNRKICGNKKNND 696
 QY 630 LCPDRVWOKKEEEMNSIKLFTKKNIO-----OSYSSNINLEFEGV--- 672
 Db 697 ECKKRIITQKDEMGKIVOHF--KTONIKGRGSDNTAELIPEDHDVLYQNLQEEFLKGD 755
 QY 673 -----FKVMDKIDKDEAKMKELMENIKRKNNEFSMLENNND-----YLENATEILL 718
 Db 756 SEDASEKSENSLDAEA-----ELKHLREIIESEDNNQASVGGVTEOKNINDKLL 809
 QY 719 DILKERTATC-----KONTNEACETSHNATINPCVYKPRGCTD--PTKNIKEIAO 766
 Db 810 NYEKDADLCLEIHEDEBEKEKGDNECIEEGENFRYNPCSGESGNKRYPIVLANKVAYO 869
 QY 767 YFKRSAYEEARNRGLHLKLGKAHEGIYKRGGRKDEKMLCLIMIKHSRNTGFSNGPCD 826
 Db 870 MHHKAKTOLASRAGSALSGLDISLAOFKNGRSTLKGQICIKINENYSNDSRNGSGGPGT 929
 QY 827 GKGTGDIOTRFVGTIEWEVEPEHMKRDEVEDYIMPRRRIHTSMLNHLQTDHPLNGNI 886
 Db 930 GK--DGHGVRRIIGTSEMNIENIGKOLSYKNVFLPRRPHMCTSLNENIDV-----GSV 982
 QY 887 V--DDLVNSFLDVLISAKYEANKIIRMYKKEKNLKGPEVTD--KHQTTICRAIRS 942
 Db 983 TKNDAASHSLGDVOLAKTDAEELIKRYKQNNI-----QLTDPYIOKQDEAMCAVAYS 1038
 QY 943 FADIDDIIRGRDLMERNGDMVLOGLHLETVENIHSKLG--KGNDKYND--ARKYTL 998
 Db 1039 FADLDIIRGRMDEDSSTDMETRLITVEFKNIKEKHGDIKDNKYTGDESCKRPAVYKL 1098
 QY 999 RENNWEANRAKWEAMKCDIKYLLKDKSGHOSFOSYSGYSDHPTPPDDVYIPOKLRMTMA 1058
 Db 1099 RADMWEANRHQVWRAMKC-----ATKGIIC--FGMPVDVYIIPORLRMTMA 1143
 QY 1059 EMYCKVQAKKEDYKLEKECKEDKNDGOGCTKESGTGCTKTECAENEYNDIITGLKMEWN 1118
 Db 1144 EMYCKAOSOEYDKIKKICADCMKSGDK--CT--OGDVDCGCKAKACDKYKEIELEKWNEDWR 1201

QY 1119 IISDKYELHEQOAMYSNGSIEASTAKNHID--RNVIETLSELYOONGK-----SNK 1171
 Db 1202 KIDSKYMLLQAKTTSTNG-----RTVLGDDDDPQOQWVDFLPIHKSIAARVLVYKA 1257
 QY 1172 SGTSDSAVIGTWTYENVAGYLHDGNFDDQSOQNEPDER-----SDGADNEKYAR 1225
 Db 1258 AGSPTEIAAARPTTPTSTAGYTHQETGIGGCOEQOPEKKNHAGASTSTTKENKETEPR 1317
 QY 1226 DKQDHDGACGCKSGSPTRVQITKKAEEKDEKTEKYNDILKENDKROYEDCJPRKN 1285
 Db 1318 QPPEVATACDCINRSOTE-----EPKKKEENVSACKIYEKILEGNGTGYGECKPKS 1373
 QY 1286 SNGYPMOC--GNINLYEDRVCMPPRRQOLCYHFLANDNEIKLOSQVNLKAPITKSA 1344
 Db 1374 ---YPMDCNNNDISHD--GACMPPRRQICLYYAHNEOTEKIKIDMLKNAFITAA 1429
 QY 1345 EPEFSWYTKSK--DGBGNELDKELKGRIPPAFLRSMTFGDYRDLFGTDSHGHRG 1403
 Db 1430 EPELSWQYKSKNDSAKILDR-----GLIPSOFLSMATTFEDYRDLCLNTDISKQNDV 1485
 QY 1404 SKLEQIDSLFKNGDOKSPNGKTRQEMWTEHSHEIWEAMICALVYIGAKKD--FTEY 1460
 Db 1486 AKAKDKIGKFPKDGSKSPSGLSRQEMWTKNGPELWKMGLALTYVYDQKRRKIKNDY 1545
 QY 1461 GYNNVFKSDK--STLLEEFKAPROFLKLTWEDDYCYTRQKILKVOEKCS--NDQLKCD 1518
 Db 1546 STDKNVQSONGNSLEEFKAPROFLKLTWEDDYCYTRQKILKVOEKCS--NDQLKCD 1518
 QY 1519 T--EONKCEYVYKMK--KKKEWIPQDKYKDEBDKRFROHIGVAVTDTGTGNADY 1574
 Db 1606 DAKHRCNQACRAYOEVVEKKEKFGSGQTNFVLKANNVQRODEYGYEYKD----- 1656
 QY 1575 LNRKFTASGCDRPGASAVYQRIOLLEKQAYDADKHCCTKFEIENDKYNISKDKC 1634
 Db 1657 -----GVOP-----IQGNEYLLQK--CDNNK--CSC----- 1678
 QY 1635 GLVKBANTGALIMQNKGPNNYNNLKEITMEDVLFPSRRLRICFHALDNGYTPREYDENG 1694
 Db 1679 -----MDGVNLSVSPREK--- 1691
 QY 1695 RKLMEVATGESYINIGYKKEKKEKIKTSDAHKYSYEVPCSAMKYSFYDLRIDILGI 1754
 Db 1692 -----PRKY-----AHKYP----- 1701
 QY 1755 DNLDEKQTEENLKKIKNKGTSYKGSJSTGPNGSTAKFPWENKECYWMAITGY 1814
 Db 1702 ----- 1701
 QY 1815 KRGDDNGSNGSARSDEDLKKGVSPPDDYPMGKNRDEGTAQOFLRWFARMGDEPCKH 1874
 Db 1702 ----- 1701
 QY 1875 EKELELVAGACNDYTCGDNEDKRRKCTDCAQYKFKISEMKPOYEKOIKRYENKDKIYS 1934
 Db 1702 -----EKC----- 1704
 QY 1935 EHPYAKDAEDAREYLDQKLIKCNKSGDCEYKCMKDYSTORLTDGSONMPASLDEPK 1994
 Db 1705 -----DC----- 1706
 QY 1995 EBEKSCNOVPRGPRVARETSPRVSILSKAFASKKEAKTAPTKQPKYENLTTEMA 2054
 Db 1707 -YQK--HVSIP-----PPPPVPO----- 1725
 QY 2055 QTRRRAAOQTRKATSTATTESDYGMVKAALISNKPDSNGGLECNPKTYGOY--PKWG 2112
 Db 1726 -----PPAPTYVYVCSIVYTLF--DTNNFSDAGLK--YGTAPASMK 1766
 QY 2113 CI-----YKSKENENGICMPRRKRLCINNIOVLNET-----EKRD 2151
 Db 1767 CIPBDTSGAGATTKGSGSDSGSICLIPRRRLVYVGLQ--EWATALPOGEAAPSRSRA 1824

```

QY 2152 NDIKEAFKCAIEPQPLMKITIE-NPAANE-----LONGT 2188
DB 1825 DLRNFAIOSAIEFEFLMDRYEEKRPOGSOALSOITSTYSDEDEPPDKLLONGK 1884
QY 2189 IPDEFRIYVYTDYKDMFEFGTDSNDKIIIVTNSVTLLENKKKODKKDELRK 2248
DB 1885 IPDEFRLMFTYLDYRDLIVHGNTSDSGNTGNSNNITVLEASGNKEDMOKIOETEQ 1944
QY 2249 I-----FWEKNKFIWEGMIGLYT-----HITDENEKIR 2280
DB 1945 ILPKNGTFLVPRKSAQTPDKMWNHAEISWKMICALYTEKNPDTLSANGDNKIEKD 2004
QY 2281 DMT-----QYN-DMTKL-----TPSLDEEYKRP 2302
DB 2005 EYVEKEFGSTADKNGTASTPTGYTKQYDEKYLKEDTSACKTPASSDPPLSDFLRP 2064
QY 2303 QFLRMTEMAEFCNRKBDLLEAGCKEY-----CN-----GSN 2339
DB 2065 PYRYLEMGONCKRRKHLAQIKHCKVEENGSGSRGCGITROYSGDEACNEMLPKN 2124
QY 2340 DGTQOE-----CAECVYQNFITKMKTEYERQREKFKKD- 2376
DB 2125 DGTVPDLKPSAKPCSSYRKWIESKGEFEKEKAYEQOKDCVNGSNKHDNGFCETLT 2184
QY 2377 -GKKYDY-----PSTERDEKATCAHEYLIMLKELCGKDCSCQKPSQPKTTQOS 2430
DB 2185 TSSKADFTKTLGPRCKPNNVEGTIPDDDKTFK-----HTKDDPCILKFSVNCCK- 2234
QY 2431 QSSDANDMPESLDYVEEFKCEPESLKKSGSIHTKITEPKIPMNCVKAAYLSKEA 2490
DB 2235 -----DECD-----NSKGTCCRNK-----NSID-----ATD1 2256
QY 2491 ENNMDTLKEKPIESTKESKSNMTNNPCDPKRPADKYIGRNPCEN----- 2543
DB 2257 ENGVDSTV-----LEWRVYASDKSGFNGD-----GLENNCRGAGJFEG1 2295
QY 2544 REENRKVDYEMKCYKSKFYQEKKRVCPREHMCRLNDELKIERLKSNTYLKMR 2603
DB 2296 RKO-----EMKC-RNVCGYV-----VCKP----- 2313
QY 2604 RTARNEGIDILKFNSENGCAMPICDITMKYSFADLGIYRGTDMLRIGGLPVEIKLY 2653
DB 2314 -----ENVNGE-----AKGHIILIOIALVR-----RWV 2336
QY 2664 KVEEYLYGKRNKRNKRNKND-----VQTFRSAMDMANRDKWATCKAPEDAKLERKG 2719
DB 2337 EYFEEDYNKIKRHSIRIKNGEISPCIKNCVEMQOKRKE-WKEITERFKD----- 2387
QY 2720 RMDGFERITLIDOKCGHNDPRVDYIPQFRMTEMSEYTCALMELEKFKSCDHCK 2779
DB 2388 -----QYKNNSDDDNVRSFLETLIPQITDAN--AKNKYIKLSKFGNSCG-CS 2432
QY 2780 TSDRCHNDYDENKCEOKTFCQYKNFVLKWLFDIOSKRYEIEQPIYTYISTDYH 2839
DB 2433 AS-----ANEKNK-----GEYKDAI-----DOMLKLKD-----KIGBEKK 2465
QY 2840 QNFVQKLTFKSCSVSESESEYIHEYSKCLINYNENOGSSNITRYAFEEETPSYKE- 2896
DB 2466 HH-----QTSDECSDTPOFOLEDD--ETLDDIETEAKKNMPPKICENVLTAQODE 2518
QY 2897 -ACSTLPSKRNPLNDCPTQONKD 2918
DB 2519 GGC---VPAENSEEPATDSGRE 2538

```

```

APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMIDIUM VIVAX
TITLE OF INVENTION: AND PLASMIDIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelisen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ. ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2182 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-487-826B-16

Query Match 9.08; Score 1743; DB 2; Length 2182;
Best Local Similarity 20.38; Pred. No. 1,4e-115;
Matches 757; Conservative 339; Mismatches 678; Indels 1960; Gaps 126;

QY 13 GNAASLEGDASKPIIKESHKS-ARNYLER-----YAKNIRHPSKYAKEHVSIGDUTK 66
DB 10 GSGGSS-SGCKGKDDSYIIVSDAKDLIDVGEKVEYKAKND--AKKTIKALKNLNT 66
QY 67 AEFRCGPSTPVNKHNYYPYPCNIDRKEHTNLRDYDVN---LRNPKH---GRQNRFE 119
DB 67 ANGRSSEYASSIE-----TCTL-VKEY-----YERYNGDGKRRHPCRKDAKNEDVNRFS 114
QY 120 DESEEG-NKIRNYKRRKNDAT-ACAPRRRRHMDKNLEANDINTQIHHLLGNVLTAK 177
DB 115 TLGGQCTYNNIKDSQOQDNVVGACAPYRRRLHLDYNEISID--TSTTHKLLLEVCAAK 172
QY 178 YEGESIYNNH--PHKGT-----SDACTALARSFADIGDIYRGIDMF-----RPNVADKY 224
DB 173 YEGNST-NHYIYQHORHNESASQICVYLARSPADIDIDYRGADLYGYONKEKEQKKL 231
QY 225 EFGLEVEYFKKIDHGMEDVAND-----YNPDG--SGNYIKLREAMWVNNRNRKVEATGCA 278
DB 232 EOKLKEDIFKKIH--KDVMTNGAQERYIDAKGGEFQLRBDWMTSNRETYMKALICHA 288
QY 279 SYKGYFMOSSESNTPLESNKGCGHKQKXVTNIDYVPOYLRWFDENGECECRNRNKKK 338
DB 289 PKRANFIKTACVNGKGTNQCCHICIGDVPFYDYVQYLRWFEWEDCCKRKKKKLEN 348
QY 339 VKDSCRNDKERLYCSHNGHDCITTIWKGILHLDNKCSTDCSKYFVWVGNQOEAFKK 398
DB 349 LQKQCRDYEQNLKCSGNGYCTKTIYKKGLVGEHGTNSVWCMRYETWIDNQKKEFLK 408

```

```

RESULT 5
US-08-487-826B-16
Sequence 16, Application US/08487826B
Patent No. 5939827
GENERAL INFORMATION:
APPLICANT: Slim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.

```

QY 399 OKETKEI-----OSYLSNDKFNINSEYKOFETKEFOYATN 441
 Db 409 OKRTEIISGGSGSKSPKTRARARSSSDN-----GESKFKYKLTKEVGOY 460
 QY 442 DFLNLMEGYCKG---GLPGEKITFNASDDKGIFTNSEYQVCPDCGVKCDIGKY 497
 Db 461 DKFLTKLKEGICQOPOVGNEKADNDFTEKTYK--TFSTETICEPCPMCGLEKGPW 519
 QY 498 THKSDNREVRNNEDEYKPPMGVPTNITVLVSGNEOGIDTCKNPNSSNTYKDNK 557
 Db 520 KVKGDJTCGSAKTYDPK---NITDIPVLTPDSOONILKTYKNEKGPAP--GGGQIKK 575
 QY 558 MECYKID-----ENINCKLEONTNINNDNPKIISPHNFELWYTYLLRDTIKWDLKT 612
 Db 576 WOCYDEHRPSSKNNNNCVEGTWDFTOGKOTVKSYNVFEFMDVHMLHDSVEMKTELK 635
 QY 613 CINTTT---HCIDECNRNCLCFDRWVKOKEEEMNSIKLFTFKKTIQOSYSNINLF 668
 Db 636 CINNNTNGNCRNNKCKTDCGCKQKVEKKQDBMAIKDHFGRQDIDVO---KGLI 690
 QY 669 EGYFFKMDKLDKDEAFKWELEMINIKRKNFEISLENNRDYLE-----NAI 714
 Db 691 VFSEYGLDLY---LKGMLTONIKDVGHDJTDIKHIKTLDEDAVAVYLGKONTTI 746
 QY 715 ELLDHLKETATICKONTNEACETSHNATNCPVPRGTOPTKNIKEIATQFKRSAYE 774
 Db 747 DKLLQHEKEQAEQCKOK--OECE-----KKAQ 773
 QY 775 EARNRGLHLKGRNHEGYKRGRRDKFNDLCRIMKHSNMLZGNSGPDCKGTGDGI 834
 Db 774 ESRGRSAE-----TRIDER-----TQDPASAGE----- 797
 QY 835 QTFVVGTEWEDVPEHMKRKHEDVIMPPRRRICTSNLEHLQTDHPLENGNITVDLVNNS 894
 Db 798 -----VEEEDDDYDEDED-----DVOVEE 820
 QY 895 FLGDVLASAKYEANKILIRMYKEKNLKGPEVTPKHOITICAIKYSFADIGDIIRGAD 954
 Db 821 EBG-----KEBGIVTEVTEVEVEET----- 842
 QY 955 LWERNDMVKLQGHLETFVGNHKSLSKRGKNDKYNDAPFKYLIRENWEANRAKWEAM 1014
 Db 843 ----- 842
 QY 1015 KCDIKYKDKSGHSTQSSYCGSDHTPLDDYIIPQKLRMTEMAEYCKVQKKEDKLKE 1074
 Db 843 -----VTE-----OE 847
 QY 1075 KCEKCKDNQOGCTKESGTGCTKTEACNEYNIIIGLKEQMIISDKYKELHQAQMS 1134
 Db 848 GVKPC-----DIVGKL----- 858
 QY 1135 VANSGEASTAKNHTDRNVIIEFLVYOONGSKNSGSTSDESAVIGTNTYENVAYL 1194
 Db 859 ----- 858
 QY 1195 HDTGNFDQOSONEFCEKSDKDEKAYAFRDKPDHDGACGCKSGSKPTRQIKTKKA 1254
 Db 859 -----ACGLKTP----- 875
 QY 1255 EEKOTECTVNDILKENGGKOVEDCHPRKNSNGYIPMOGGINLV-----EDPRVCM 1308
 Db 876 -----GGEK-----FPNMKCVTPSGVSTATSGKQALICVP 906
 QY 1309 PRKOKLYHFLA---NDNEIKKLOQV-----NIKEAFIKAAAEFTFSWYTY 1353
 Db 907 PRRRLLYGGISLQMSRSGDETTVESSSEATAPSOSESEKRLTAFTIESALETFFLMHY 966
 QY 1354 K-----SKDG-----EGNELDKELKE--GRIPAPAFRSMYTYGDRDFLFG 1393
 Db 967 KEKKPRATOGAGLGVSLPEPSPGEDPQTOQOTGIVTIPDFLRQMFYITLADYKDLI-- 1024
 QY 1394 TDISGHEGSKLKEQIDLSLEKNDQKSPNGKTRQEWMTESHSHEIWMALCALVIGAKK 1453

Db 1025 ----- 1024
 QY 1454 DFEFNYGNNAVFSDKSTLLEFAKRPQFLRMLEWMDYCYTRKYLAKDQEKCKSND 1513
 Db 1025 ----- 1024
 QY 1514 QLKCDTBCNKCEDVYKYMKKKKKEMIPDKYKDERDKRPRDROHIGVMTDTGTNAD 1573
 Db 1025 -----YSSNDIS 1032
 QY 1574 YLNRKFTASGDKPGSAVVOIRNIOLEKQATYDADKHCCGCTKFIENDDKYINISSKDC 1633
 Db 1033 -----DTGKQTPSSSNNL 1047
 QY 1634 KGLVKEANTGAIKWNKGNNYNNIKELTEDVLFPSRLICFHALDNGYTPPEKDENG 1693
 Db 1048 KNIVLEAS----- 1055
 QY 1694 LKRRLMEVATEGYNLGOYKKEKKEKIKTSDAKYSTEVPKCSAMKYSFYDLRDIILG 1753
 Db 1056 -----GSTE-----OEKEMK----- 1066
 QY 1754 IDNLEDEKOTENLKKIFNKGSTVGKSGDSTGNPGSTARKPFYNNKKECVNNAMTCG 1813
 Db 1067 -----QIQAKIKKILN--GATS--GVPPYTNYSYKTPQOTWENIADIMWAMCA 1113
 QY 1814 --YKGRDGNNGSNGARSDEDLKCGSVSPDDYPMGRNDEGAYOPLRFAEWGEDFC 1871
 Db 1114 LTYKE-----NDAR-----GTS----- 1125
 QY 1872 KHEKELEKLVACADYTCGDNEDKRRKCTDACIQYKFFISEMKPQYKQIKKYGENDK 1931
 Db 1126 -----AKLEONKD- 1133
 QY 1932 IYSEHPVAKDAEDAREYLDKOLKICENKSGDCYKMKDVSOTRLTDGNSONNPASLDD 1991
 Db 1134 -----LKK-----ALWDEANKNP----- 1147
 QY 1992 EPKEVEGKNGOVRGPPRVARRETPSPRVSLSKATASKKEAKTAPPTKPKVENLTTE 2051
 Db 1148 ----- 1147
 QY 2052 MRAOTRRAAQOTRKRTSTATTESDVTGWKAILNSKPSRSGIGCNPKITGYQPKW 2111
 Db 1148 ----- 1147
 QY 2112 GCITYGSKENENGICMPRRKLCINNIOYLYNETENKROUNDIKAFIKCAIETOFPLML 2171
 Db 1148 -----IEKYQYNNVLE----- 1159
 QY 2172 KYIENPAENELONGTIPBEFKRIMYTYGVDYDMFEGFDISNDKIITYVTSVTTILN 2231
 Db 1160 -----D 1160
 QY 2232 ENNKKODKKKDEBLRKIEWEKNKFIWEGMITYGLIYHLDENEKERINDYQYNDMTKL 2291
 Db 1161 ESGAKSND-----TIQ 1171
 QY 2292 TPSLEEVKRRPOFLMPTTEMAEERONKRRQOLKLEAGC-----KEYCNGSNDK--- 2342
 Db 1172 PPLTAKNFVEITPFRMLHEGNSCFERARLQIHECDEDEGEKQY-----SGDEYCE 1227
 QY 2343 -----TOCEACVYTONFIKWKTEYEROREKFKDKDGKYYDPTSTED 2389
 Db 1228 EIPSKOYNVLQDSSGAKRCRLKTIWIEKKTEYKQKAYEQK-----SYTEBORD 1282
 QY 2390 IEKATCAHEYLANKLKELCGNKDCSCMQRSSQLPPTTOOSOSDANDMPESLDYPP--E 2447
 Db 1283 -----KC-----QOSNNNANERSRLGASPTA 1306
 QY 2448 EFNKCEPBLSKKSMIHTKITEPIPMCVKAKAYIYLSKEENMDITLKEKPIPIES 2507

```

DB 1307 EF-----LQKLS-----CKNDNGT-----ENGDNKIDK-NPDKT 1337
QY 2508 TKESKNSWTNNPCDPKPPYAPDKYIGRRNCPENRENRFYVDEWKCYKNSKFEYQK 2567
DB 1338 FKEAHS-----CDP-CPITGVK-----CQNGH----- 1358
QY 2568 KRYCVPRRRHMLRINDEKIRLKDSDNYLKMVRTANNEGIDIKNSENCGCAMP 2627
DB 1359 ---CVSANGKECKNN--KTADIKN-----KTDNGNIEMVSDSTN----- 1398
QY 2628 ICDTKYSPFADIGYGTMLRIGLYLPVEIKLYKFEYIGKMRANKKGRNKYDVO 2687
DB 1399 -----TPEHLCD-----CKSSGIFPGI----- 1415
QY 2688 TFRSAMWANDARKDWMK-AMTCKAPEDAKTERKGRMDGFERITLID-KCGHKDDPVDYD 2745
DB 1416 -----RKDEMCANVC-----GVDICTLEKKIKNOGEB---KRY 1447
QY 2746 IPORFRMWTESYCYKALMEELKFRKSCDHC-KTSDCKNDYDENKCEQCTRCQYK 2804
DB 1448 ITMK-ELKRWLEFYF-----LEDYNRIKKIKLCTKEDGCK-----CJGCTE-- 1490
QY 2805 NEFLKXKSLFDISNKKYKELEPIYTKISTYHQNQVQKLTFRKSECVSESEYLHE 2864
DB 1491 ---KM-----VO---EKTEWO----- 1501
QY 2865 TSCKLNYKFNENDSNIPTAYEETPKSYKEACSCITLPSKNPLDNCPTQNDGCKELQ 2924
DB 1502 -----KIND-----TY-----LE 1509
QY 2925 TTFPCSKNDYDNNLDMNNAVLYNSDDNGVLIIPRRHRLCTRPITAYVRKGDKEILK 2984
DB 1510 QY-----KNDGDNLTJNF-----LEOFYR----- 1529
QY 2985 KLLTSAFSGQLGQYKSEELCFEAMKSYADSDIITKGTMDMTLSSEIKIKIFEET 3044
DB 1530 -----TFKNAIKPCDGLD----- 1543
QY 3045 SNEATENRKTWMENNRQIWHAMLCGYKATSKYTLDEGWCOLPKDETNQFLRWLIEMA 3104
DB 1544 ----- 1545
QY 3105 KQCKEKKAHVSDSLKTCRPSNEDNFEASBLAQPCQCONDIRYISLILIKKTMENLNT 3164
DB 1546 -----KTSQGINSTDN-----SONGNNDN-----VYLCLLNKLOK 1575
QY 3165 KYKOLDOSSGNIDNKPSEENVQSYKSDOCALENDINETYTKNNENN----- 3217
DB 1576 KISECKEONHSGQYQF-PCD---NSSLSGKSTLVEDDYDE---QNPENKVEQPKFC 1626
QY 3218 -EFKEVYLKLLPGLYVEDETHRNHVLGNI--KEEQYVRPRALYFTPHVDSFYQAPL 3274
DB 1627 PDKKEPKKENDDEVGTCGGEKKYEDSVIEQKEEBAASAPESPPLTF-----EAPK 1680
QY 3275 FSTHRAVQYD-----KN-----DIAKSSISVIYALGL---IALHFKKKFKFS 3316
DB 1681 KEENVVYKPPPKRRKRIKTRNLDPHAPVLPALMSSTIMMSIGIFEAFTFYKLKTKTKS 1740
QY 3317 SV-DLRIILNIPQGEYMGPLESKNRIPIYRSGPYGKTYIYWGEGDSSGDE-KFAEMSDT 1799
DB 1741 SVNLRLQIILQIPASDIPLKSSNRKIPYASDRHKKTYIYWGEGDSSGDE-KFAEMSDT 1799
QY 3376 SDITSSSEYEELDINDIYVPGSPKTYLLEVLEPSKRDIPSD-----DTPSNTPR 3428
DB 1800 TDTTSSSEYEELDINDIYVPGSPKTYLLEVLEPSKRDITQNDIHDSDIPNSTPTP 1659
QY 3429 TNPFDIDEMNELKHDVYSQYL---PTEPNNTKASADIPMTEPNITYSNPEKPPITIS 3485
DB 1860 P--ITDEMNQOLKDFISNMLQONTQTEPNITLHDNDV--NNTHP-TMSRNHMQKPFIMS 1914
QY 3486 IHRDLYTGKEISY-----NINMSTN-----TNDIPMNANDSGYRGIDLINDS 3529
DB 1915 IHRNRLSGEELYNDMFENSGNPNINISDSTNSMDSLTNSNHSFYNDKNDLYSGIDLINDA 1974

```

```

QY 3530 LVVLNLLIY--MMK 3541
DB 1975 LSGNHIDIDYDMLK 1988

RESULT 6
US-08-568-459A-10
; Sequence 10, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xia-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,459A
; FILING DATE: 07-Dec-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-568-459A-10

Query Match 5.3%; Score 1032.5; DB 2; Length 700;
Best Local Similarity 34.8%; Pred. No. 2e-65;
Matches 247; Conservative 110; Mismatches 250; Indels 103; Gaps 22;

QY 136 NDATACAPRRRRHMCNLEALNDINTONTIHDILGNLVYAKYEGESIVNNH--PHKGT- 192
DB 5 NKVGCACAPRYRLHLCDNLSEID--TTSSTHKLLEVCMAKAYGNSI-WTHYQHORNTN 61
QY 193 ---SDACTALAREFADIGIVRGIDMF-----KRVVHDKVEGTGLREVRKTHIDGMEDE 242
DB 62 EDSASQICTVLANSFADIGIVRGKIDLYLGDNKEKQKRLKLDIFKTH--KDV 118
QY 243 VKND-----YNPDG-SGNYKYLREAWNVNRNKVWEAITCDASYSGYFMQSESNTPLFS 296
DB 119 MKTGAQERYTIDAKGSDFPQLREDWMTSNRETYWKAALICHAPREANYFIKTACNVGKGT 178
QY 297 NPKCGHOGQVPTNLDYVPOYLRWFDEMGEBFCFKRNIKLAKVADSGRNDKERLYCSHG 356

```



```

Db 179 NGQCHIGDVPYTPDYVPQYLKMFEEAEDPCRRKKKLENLQKCHDYBQNLGCSGNG 238
QY 357 HDCTTTIRKKGLHLNDKCTDCTCKAYEYEWLGNQOAFKROKREYKEI----- 407
Db 239 YDCTKIYKKGLVIGECTNCSVMCRMYETWIDNQKKEFLKOKRKRYETETISGGSGSKSP 298
QY 408 -----QSYLSDNKNKVVNNINSYKYQFYEKLEKETOYATNDFLNLNGKYCKG--- 456
Db 299 KRTKRAARSSSSSDNN-----GYESKFYKKLEWYQDVDFELKILNKEGICOKOPQ 350
QY 457 -GLPGEKDITFTNSADKGIYFSEYCOVCPDGCVKCGIKYTHKSDNRERANNEDYKP 515
Db 351 VGNKEADNVDTNEKYK-TFSRFEICEPCWGLEKGGPWYKCGKTCGSAKTYIDP 409
QY 516 PWGKAPNITVLYSGNEQDITOKLEFNCNSSTNYKDKNNQKWCYKYD-----ENINRC 570
Db 410 K---NIDIPALYDKSOQNLKRYKNCEKAP-GGGQJIKKQCYIDEHRPSSKNNNC 465
QY 571 KLEQNTENNDNPKIISFHNFEFLWYTYLLRDTIKNDKLTCTINNTT-----HCIDECN 626
Db 466 VEGTWDKFTQOKQVKSYNFFWDVWMDLHDSYEMKTELSCINNTNGTCRNNKCK 525
QY 627 RNCLEPFRWYKQKEEENSIKLTETKKNNIQOQSYYSNNLFGYFPKVDKLDKDEAKY 686
Db 526 TDCGCFQKWEKKQOEMMAIKDHGKOTDIYQO---KGLYFSPYGVLDV---LKG 576
QY 687 KLEMNIRKKRKNESNLENNRDYLE-----NAIELLDHLKETATICKDNN 732
Db 577 GNLQNTKDVHGTDIDKIKHLKLEDEDAVAVYLGKDNNTTIDKLQHEKEQAEQCKOK- 635
QY 733 TNACE-----TSHNATNPGVKPRGTOPTKNIKEIAQYKRSAYEE 775
Db 636 -QECEKKAQOESGRSAETREDEKQ---QPADSAGEVEEEDDDDYDE 681

RESULT 7
US-08-487-826B-10
; Sequence 10, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Slim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelien, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH21.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 10:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-487-826B-10

Query Match 5.3%; Score 1032.5; DB 2; Length 700;
Best Local Similarity 34.8%; Pred. No. 2e-65;
Matches 247; Conservative 110; Mismatches 250; Indels 103; Gaps 22;

QY 136 NDALACAPRRRRHMDCKNLALNDINTQNIHDLGNVLTAYEGESYVNNH--PHKGT- 192
Db 5 NKVACAPYRRLHLCDYNEISID--TSTTHKLLLEVCAAKYEGNSI-NHYTOHQTN 61
QY 193 ----SDACTALRSPADIGDIYRGIDM-----KPNVHDKYETGLREYFKKIHGMEDE 242
Db 62 EDSASQCLTVLARSFADIGDIDYRGDLVLYGDNKEQKRLKLEOKLIDIFKIH--KDV 118
QY 243 VRND-----YNPDG--SGNYKLEAMNNVNRKMYEATITCDASYSGVFMQSESNTPLFS 296
Db 119 MKNGAQERYIDDAKGDFQFQLRDMWTSNRETWAKALICAPKREANFTITACVKGKT 178
QY 297 NPKCGHKGQVPTNDYVPQYLKMFEDWEGEEFCRRKNIKLVKQVDSNDRERLYCSHNG 356
Db 179 NGQCHIGDVPYTPDYVPQYLKMFEEAEDPCRRKKKLENLQKCHDYBQNLGCSGNG 238
QY 357 HDCTTTIRKKGLHLNDKCTDCTCKAYEYEWLGNQOAFKROKREYKEI----- 407
Db 239 YDCTKIYKKGLVIGECTNCSVMCRMYETWIDNQKKEFLKOKRKRYETETISGGSGSKSP 298
QY 408 -----QSYLSDNKNKVVNNINSYKYQFYEKLEKETOYATNDFLNLNGKYCKG--- 456
Db 299 KRTKRAARSSSSSDNN-----GYESKFYKKLEWYQDVDFELKILNKEGICOKOPQ 350
QY 457 -GLPGEKDITFTNSADKGIYFSEYCOVCPDGCVKCGIKYTHKSDNRERANNEDYKP 515
Db 351 VGNKEADNVDTNEKYK-TFSRFEICEPCWGLEKGGPWYKCGKTCGSAKTYIDP 409
QY 516 PWGKAPNITVLYSGNEQDITOKLEFNCNSSTNYKDKNNQKWCYKYD-----ENINRC 570
Db 410 K---NIDIPALYDKSOQNLKRYKNCEKAP-GGGQJIKKQCYIDEHRPSSKNNNC 465
QY 571 KLEQNTENNDNPKIISFHNFEFLWYTYLLRDTIKNDKLTCTINNTT-----HCIDECN 626
Db 466 VEGTWDKFTQOKQVKSYNFFWDVWMDLHDSYEMKTELSCINNTNGTCRNNKCK 525
QY 627 RNCLEPFRWYKQKEEENSIKLTETKKNNIQOQSYYSNNLFGYFPKVDKLDKDEAKY 686
Db 526 TDCGCFQKWEKKQOEMMAIKDHGKOTDIYQO---KGLYFSPYGVLDV---LKG 576
QY 687 KLEMNIRKKRKNESNLENNRDYLE-----NAIELLDHLKETATICKDNN 732
Db 577 GNLQNTKDVHGTDIDKIKHLKLEDEDAVAVYLGKDNNTTIDKLQHEKEQAEQCKOK- 635
QY 733 TNACE-----TSHNATNPGVKPRGTOPTKNIKEIAQYKRSAYEE 775
Db 636 -QECEKKAQOESGRSAETREDEKQ---QPADSAGEVEEEDDDDYDE 681

RESULT 8
US-09-210-288-10
; Sequence 10, Application US/09210288
; Patent No. 6392026
; GENERAL INFORMATION:
; APPLICANT: Slim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.

```



```

QY 1419 QKSPNG-KTROEWMTESHEIMEAMCAL-----VKIGAKKDDPTENYGYNNYKF-S 1468
DB 547 YKISNSIRYKRSWMTNGPIYMEGMCALSYDTSLNNVNPETHKRLTEGNNNEKVIJFGS 606
QY 1469 DKSTTLEEFARPOFLWLTWYDDYCYTRQKYLKDVOECKSNDOLKCTECCNK---C 1525
DB 607 DSSTTTSKFSERQFLWLTWYDDYCYTRQKYLKDVOECKSNDOLKCTECCNK---C 1525
QY 1526 EDYVK-----YMKKKKEWIPQDK--YKDERDKKRPD--RQHIGVWVTDYTG 1569
DB 666 KDCKQYHWSWIGWIDYKQKGRYETVKKIPLYKEDKDVNSDDADYLTQLONNKCV 725
QY 1570 NADYDYL-----RKFTASGDKRPGSASVYQRIQLLEKQAYYDADKHCCTCFIENDK 1623
DB 726 NGTDECEYKCMKHTSSTNSDMPES-----LDERPEVKDK-CNC---VPNECN 771
QY 1624 YTNISS-----KDKCKGL 1636
DB 772 ALSVSGSGFPDGOAFGGVLEGTCKGL 798

RESULT 11
US-09-210-288-8
Sequence 8, Application US/09210288
Patent No. 639026
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chltnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210, 288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH21.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-8

Query Match 3.3%; Score 633; DB 4; Length 921;
Best Local Similarity 27.8%; Pred. 1.4e-36;
Matches 224; Conservative 113; Mismatches 264; Indels 206; Gaps 42;

```

```

QY 948 DIIRGRDLERNQDMVKLOCHLETVEGNIHSLKCKGNKYN-----DAPKYLKLENN 1002
DB 80 DYTASDI-ENGNSI-----GNIDMVYSDKANGFNGLDAGSANIFGIRKEQ 128
QY 1003 WEANRAVYAMACDIYLLDKSGHOSTQSSYCSHTFLDYIPKLLMNTMAWYC 1062
DB 129 WKC-AKV-----CGLDVCGIKNGN-----GSIDKDKOJIIIRAL--LKRWVYFL 171
QY 1063 KVQKEEYDLKLECKECKEDKNDNGGCTKESGCTGCTKTEACNEYNIDITGLKEDMNTISD 1122
DB 172 -----EDYKNINAKISHTCKRDNSTCTINDCPNCT-CVE-----EWINDKREBKAIK 220
QY 1123 KYKELNE-----QAQMSVSN-----SGIEA-----SSTAKNH 1149
DB 221 HYQTQENNGDNMKSILVTIDILGALOPSDVYKAIKPCSGILAFESFGGLNADNSEKKEG 280
QY 1150 ID-----RNYEFSELYQONGSKNSGTSDESAVIGTNTYNNVAGYLHDTGNFD 1201
DB 281 EDYDLVICMLKNEKQIOECKKRRHGETSVENG--GKSTPLDNTTLEE-----E 327
QY 1202 DCOSONE-----FCDEKS--DGKDNK-----YAPDRKPDHDGACGCKSGSKPRVY 1247
DB 328 PIBEENQVEAPNICQKQVEDKCKKEEBEETCTPASVPEKPVH--YAKRRTFTPEVVK 385
QY 1248 IKTKKAEEKDTECKTVNDILKENDGRQVEDCHPKKNSNGYPDQGNINL-VEDEPRVC 1306
DB 386 I-----WRGRNRKTTCEIYVAEMLKDKNGRTVGEYRKT--YSEWTCDESKIKMGHGAC 439
QY 1307 MPRRQKLYHFLANDIEIKKIOSVN-LKAPFKSAATFTFSWY-YYSKQCEGELD 1364
DB 440 IPPRRQKLYHFLANDIEIKKIOSVN-LKAPFKSAATFTFSWY-YYSKQCEGELD 1364
QY 1365 KELKEGKIPPAFLSMETFGDYRDLFGTDI-----SGHGEGSKLKQDISLFXNGD 1418
DB 494 EKLKGLIIPEDFKRQMTYTTADYRDLGLGDISKKTSGVCG--KVCNIDVF---- 546
QY 1419 QKSPNG-KTROEWMTESHEIMEAMCAL-----VKIGAKKDDPTENYGYNNYKF-S 1468
DB 547 YKISNSIRYKRSWMTNGPIYMEGMCALSYDTSLNNVNPETHKRLTEGNNNEKVIJFGS 606
QY 1469 DKSTTLEEFARPOFLWLTWYDDYCYTRQKYLKDVOECKSNDOLKCTECCNK---C 1525
DB 607 DSSTTTSKFSERQFLWLTWYDDYCYTRQKYLKDVOECKSNDOLKCTECCNK---C 1525
QY 1526 EDYVK-----YMKKKKEWIPQDK--YKDERDKKRPD--RQHIGVWVTDYTG 1569
DB 666 KDCKQYHWSWIGWIDYKQKGRYETVKKIPLYKEDKDVNSDDADYLTQLONNKCV 725
QY 1570 NADYDYL-----RKFTASGDKRPGSASVYQRIQLLEKQAYYDADKHCCTCFIENDK 1623
DB 726 NGTDECEYKCMKHTSSTNSDMPES-----LDERPEVKDK-CNC---VPNECN 771
QY 1624 YTNISS-----KDKCKGL 1636
DB 772 ALSVSGSGFPDGOAFGGVLEGTCKGL 798

RESULT 12
US-08-568-459A-4
Sequence 4, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chltnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:

```

```

ADDRESS: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-0176
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-4

Query Match 3.2%; Score 617; DB 2; Length 1435;
Best Local Similarity 21.8%; Pred. No. 3,9e-35;
Matches 332; Conservative 215; Mismatches 557; Indels 418; Gaps 69;

QY 1101 EACNEYNDITGLMKED--WNTISDKYKELHQAQMSVNSGSIASSTAKNHIDRNVIEF 1157
DB 20 KARNET-DI---KENKELDYKKEKFNELDKKKGYNVQKT-----DKKIFEF 62
QY 1158 L-SELYOONGKSGNS---GTSDESAYVIGNTTYENVAIYLDHGNFDDCOSQNEFCE 1212
DB 63 INKIDLINNSKFNKRWKSYGTPDN---IDKMSLIN----- 96
QY 1213 KSDGDNKRYARDRPODHGACGCKSGSKPYRVO-ITKKAKEEDIECKTIVNDILKEN 1271
DB 97 ---KHNEEMFNNNYQSPFSLTSSLIKQNKYVPINAVYRSLSLFDSR---INNGRNTS 149
QY 1272 DGKQVDECHPRKKNNGYDPMQCGNINLYEDPRVCMPPRRQKLCYHFLANDNEIKKLSQ 1331
DB 150 SNNEVLSNCRERKRG--MKWDCKRN-DRSNVYCIPIPRRIOLCTVNS---IKITYK 201
QY 1332 VNLKFAFIKSAAEFFSWYKYKSKDGGNLDKELKGGKIPAPFLRSFVYEGYRPL 1391
DB 202 EIMKHFIIEASKKE---SOLLKAKNDKNYS-----KCNDKLSFIDYGHIA 246
QY 1392 FGTDSIKHGEGSKLEQIDSLFKNGDKSPNGKT---ROEWTEHSEHIEWEAMLCALVK 1448
DB 247 MGNDDPG-GYSTKAENKIQEYFGAHGEISEHKIKNFRKEMWNNFRKIKWAML----- 300
QY 1449 IGAKRDDTEYGNVNVFSDKSTLLEFPAKRPQFLMLTEYDYDYCYTRQYKLVQEK 1508
DB 301 ---SEKKNIN-NCKNIPQEEI---QITOWIKEHGHEFLERDRSRSLPKRSK 345
QY 1509 CKSNDOLK-CDECKKCKCEDYVYK-MKKKKEMIPQDKYXKDERDKRRDRHIGWVVDY 1566
DB 346 CKNNITLACEKECIDLPCMKYTRDMITRSKFEMHTISKEFETQKYKE----- 392
QY 1567 TGTNATDYLNRKFTASCGDKPGSASVYVORNIOLLEKQAYYADKHCGC---TKFIE--- 1619
DB 393 ---NAENYL-IKISEKNKD-----AKVSLILNNDADAEYSKCDCKHTTTLVKSIVLN 439

```

```

QY 1620 -ND---DKYTNISSKDKCK-GLVKEA-NTGALKQNKGNPNVNNIKELTEDVLPSPRL 1672
DB 440 GNDNTIKERREHIDLDDESKFGDKNSVDTNKTWKECKNPYLLS-----TKDVCVPRRO 494
QY 1673 RICEFALDGNVTDPEYKDENGL--RRKLMEVATEGYNIGQYKKEKEKIKTSDAHKY 1730
DB 495 EICLIGNDRIT-----DKNLMIKHEHIALAIYISRLIKRKYKKNDDKE----- 538
QY 1731 STEVPPCSAMKSFYDLRIIIGIDNLEDEKQTEENLKKIKNKGTSYKGSSTGTGNP 1790
DB 539 ---VCKLIINTFADIRIDIGITDYWD---LSNKLKGLINTSKYVHRKKKN----- 585
QY 1791 GSTARKEFWNENKECVWMAIMIGYKRGDRDNGSNGARSDEDLKKGGSVPDDDYPMGN 1850
DB 586 DKLFRDEMMAKVIKKDVMVNIYSVK-----DITYCKE 617
QY 1851 RDEGTAQFLRPAEWGDEFCHEKELEKIVAGCNDYTGDNEDKRRKCTDACTQYKFF 1910
DB 618 DDIENIPQFFFWFSEWGDYCODTKIMLETLLKVECKEKPCEDD---NCKSKNSYKEM 672
QY 1911 ISEMKPOYEKOIKKYE---NKDKIYSHPAKDAEDAREYLDQKIKCENKSGDCDY 1966
DB 673 ISKKEEYKOKAOYOEYOKGNMYKYSFKSIKREYVYLKYSSEK---CSNLFEDDF 727
QY 1967 K-''---C--MKDYSTORLTDGNSONMPASLDDEPEYK---EGKCNCQVPRG 2007
DB 728 KEELSDYKKNKCTMCPYKVDVPI-S-IINNEDTSEANPEENTETLAHTEPISIEGKG 786
QY 2008 PPRVRETPSPRVSILSKATASKAKTAAPTQOKRYVENLTENRAOTRRRAAQOTRK 2067
DB 787 NQOKRRDSDS---LSKISVS-----PENSRL-----ETDAK 814
QY 2068 RSTATTESDVG--DMVAILSNKPSDGRGIEGCPKTYGQYPRKGCIVGSKENENIC 2126
DB 815 DTSNLKLKGDVDISMPRAVIGSSPN---DNIN-----VTROGDNISGV- 855
QY 2127 MPRRKLCLINNIQYLVNTEENKRDNDIKFAIFKCAIETQFLMLKITYIENPAENELON 2186
DB 856 ---NSKPLSDVDRDKLEQNDSESEYV-----NHISKPSITNGDS 899
QY 2187 GTIPPEFRKIMTYIGDYKDMFFGTDISNDKKIITVNSVTTILNENKKODKKKDEEL 2246
DB 900 GS-----GSATVSESSSSNTGLSIDDRNGDFVFTQDANTEDEVIRKENADKDEDE 951
QY 2247 RFIWEKKNKFIWEGMIGLYHLDEN---EKKEIDNTQYNDMTLFTSLEFYRRP 2302
DB 952 KGADER---HSTSELSPEEKMLTDNGNSLN-----HEVAKH 990
QY 2303 QFLRWFTEMAEBFCNKRREOLK-----LEAGCKEYBCNGSN----- 2339
DB 991 TSNSDNVQSGGIYVMNYEKELDTLENPSSSLDEGKAHELSERNLSSDQMSNTPPL 1050
QY 2340 DGKQOECAEACVYONFIFKKKTEYEROREKPKD-----KDGAKKYDPPSTER 2388
DB 1051 DNTSEETTER---ISNNEKYVEREDERTLTREYEDIVLKSNNRESDDGELYDE---NS 1104
QY 2389 DIEKATCAHEYLNNKLELCNKCQSCQKPPSSOLPTQOQSQSDANDMESID---Y 2444
DB 1105 DLSTVNDSEDAEAKMK---GN-DTSEKSHNSOHIEEDQK-----NDMKTVDGLGTH 1155
QY 2445 VPEEFNKCEPELSK---KGSMIHTKKTIEPKIIPANCYKAAAYLSKEAENNDITYL 2499
DB 1156 VONEISVYVTEIDEKLERESKSIH--KAEEHLSHTDIHK-----INPEDRNSNTLHLK 1209
QY 2500 EKFTPIESTKEKESKNSVTNN 2521
DB 1210 D-----IRNENERHLTNON 1224

```

RESULT 13
US-08-487-826B-4
; Sequence 4, Application US/08487826B

Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Slim, Kim L.
APPLICANT: Chilton, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-Zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPL
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-8550
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-487-826B-4
Query Match 3.2%; Score 617; DB 2; Length 1435;
Best Local Similarity 21.8%; Pred. No. 3.9e-35;
Matches 332; Conservative 215; Mismatches 557; Indels 418; Gaps 69;
1101 EACNEVDITGLMKEQ--WNIISDKYKELHEQAQMSVNSGIEASSTARNHIDRVNIEF 1157
20 KAREY-DI---KENEFIDYVKEKEFNEIDKKKKYGNVQKT-----DKRIFTF 62
1158 L-SELVQONGSKSKS-----GTSDSAVIGTNTTYENVAGVLYDGTGFPDDCQSQNECDE 1212
63 IENKLDILNNSKFNKRSKSTGTPDN---IDKNSMLN----- 96
1213 KSDGKDNKAFBDKPODHGACGCGSGSGSPTRVQ-ITKKAKEKDETECKTVNDILKEN 1271
97 ----KHNNEMFNNNYVSFLSTSLIKONTYVPLNAVRSRILSLFSDR---INNGANTS 149
1272 DGKQVEDCHPKKNSNGYPRMOCNINLVEDPRVCMPPROKLCVHFLANDNEIKLQSQ 1331
150 SNNEVLNSCKEKRKG--MMWDCRKN-DRSNVYCIPIRRIOLCIVMS---TIKITYTK 201
1332 VNLEAFIKSAATPFPSWYYSKDEGNEHDEKELKEKIPAFILSMYTFEDYDPL 1391
202 ETMDQHEITLSKKE---SOLLKKNKYNK-----KFCNDLKNSEFLDYGHLA 246
1392 FGTDISGHEGSKRLKEQIDSLFENGQDKSPNGKT---ROEMWTEHSHEIWEAMICALVK 1448
247 MGNMDPG-GYSTKAENKIDGVFGANGELISEHKIKNFRKEMWEPFEKLMLEAML----- 300

1449 IGAKKDDFTENYGYNNYKFSDKSTLLEFAKRPQFLMLTWMYDDYCYTRQKYLKVOEK 1508
301 -----SEHKNNIN-NCKNIPQEL-----QITQWIKEMGEFLEEDNSKLPKSK 345
1509 CKNSDOLK-CDTECNKKCEYVVK-MKKKEWLPDQKYDERDKKFPDQHGIVNVTDY 1566
346 KNNITLVEACEKCIDPCMRYRMIIRSKFEWHTLSKEVEYQYPKK----- 392
1567 TGNATDYLNRKFTFASCGDRGASVYQRNIOLLEQAYVADAKHCC---TKFIE--- 1619
393 ---NAEYTL-IKISEKND-----AAVSLINNCDAEYKYCKCKKTTLLVKSVLN 439
1620 -ND---DKYTNISSKDKCK-GLVKEA-WTGAIRKQKGNPNYNNLKELEDEYLFPSRL 1672
440 GNDITIKEREHIDLDGFSKFGCDKNVDTNTKWECKNPYILS---TKDYCVPRRQ 494
1673 RICEHALDGNVTPENVDEGL--RKLMBVAATEGYNLGOYKKEKKEKITSDAHKY 1730
495 ELICGNIDRIY---DKNLMKKEHILAIYESHILKRYKKNKDKK----- 538
1731 SYEVPSCAKKSYFYDLRIILGIDNLEDEKQKTEEMIKTFNKNGTSGVSGSDSTGNP 1790
539 ----VKIINKTPADIRDIIGTDYND---LSNRKLVKINTNSKYVRNKNK----- 585
1791 GSTARKFEFNNENKECVNNAMICGYKRGDRDNGNSARSDEDLKCGSVSDDDYPMGN 1850
586 DKLFREDMWVYIKKDVNVNLSWVK-----DKTYCKE 617
1851 RDEGTAVQFLRWAPGEGEDCKKKELEKLVGACNDYTGCDNEDKCKCTDACTQYKFF 1910
618 DDIENIPQFFRWFSWEDDDYCODTKMIEFLKVECKKPEED---NKSCKNSGSEY 672
1911 ISEWKPOYKQIKRYGE---NKDIYSEHPVAKDADAREYLDKOLKICENSGDCEY 1966
673 ISKKEEYNQAKOYQYQGNNTKMTSEKSIKPEYLLKYSK-----CSNLNFEDEF 727
1967 K-----C-MKDVSTORLTDGNSQNPASLDEPKVEY---BGKNCQVPRG 2007
728 KEELSDYKKNKCTCPFKVDYPIIS-IRNNEQTSQEAIVPEENTEIAHRTPEPISSEPKG 766
2008 PPRVRRETPSRVSLISKAKASKREAKTAPPTOPKVENLTTMRQOTRRRAOQTRK 2067
787 NEQKERDDSD---LSKISVS-----PENSRP-----ETDAK 814
2068 RTSTATTTESDVG-TWYKAILSNKPDGSGIEGCPRTYQYPRKGVKSGKENEGIC 2126
815 DTSMLLKLKGDVDSMKAVIGSSPN-----DNIN-----VTEQGNISGV- 855
2127 MPRRKRKLCINNIQYLYETENKRDNDIKAEFIKCAAIETQFLMKYTIENPAEENELON 2186
856 ---NSRPLSDVDVPRDKKELEDONSDESEYV-----NHISKSPSINNDDS 899
2187 GTTDFEKRIMYIYGYDKMFEGTDISNDKITTIVNSVTTILNENKKKOKKKKEEL 2246
900 GS-----GSATVSSSSSNTGLSIDDDRGDFVYRTDTANTDEVIRKENNDKODE 951
2247 RKIFMERKKFIWEGMITYGLYHLTDEN---EKEKIRDYQYQYDMTKLTPSLEEFYKRP 2302
952 KGADDER-----HSTSESLSPEKMLTLDNGGSLN-----HEBYKEH 990
2303 QFLRWETMAEEPCNKKREKDLK-----LEAGKEYECNGSN----- 2339
2340 DGKTOCEACACVYQNFIFKMKTYERORERKFD-----XDGKKYKQYPESTER 2388
1051 DNTSEETTER--ISNNYKVNREDERLTKEYEDYILKSHMARESDGELYDE---NS 1104
2389 DIERKATCAHEYLNNKLKELCGNKDSCMORPSSQLPTTQDSOSSDANDMPESID---Y 2444
1105 DLSIVNDESEDAEAKMK---GN-DTSEMSHNSQHISSDQK-----NDMKYVDDLTGTH 1155

QY 2445 VPEENKCEPELCK-----KSMITHKKTPEKIPANCVKAAVYLSKEAENNDITLK 2499
Db 1156 VONEISVPTGEIDEIKRESEKSIH--KAEERLSHTDIK-----INPEDRNSNTLHLK 1209
QY 2500 EKFIESTKEKESKSNWTNN 2521
Db 1210 D-----IRNENRHLTNN 1224

RESULT 14
US-09-210-288-4
Sequence 4, Application US/09210288
Patent No. 6392026
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chluis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-4

Query Match 3.2%; Score 617; DB 4: Length 1435;
Best Local Similarity 21.8%; Pred. No. 3.9e-35;
Matches 332; Conservative 215; Mismatches 557; Indels 418; Gaps 69;

QY 1101 EACNEYNDITGLMKEQ--WNITSDKRELHQAQMSVNSGIEASSTAKNHIDBNVIEF 1157
Db 20 KARNREY-DI---KENEFKLDVYKEKFNELDKKKYGNQKT-----DKITTF 62
QY 1158 L-SELYQONGGSKNS---GTSDESANVIGINTYENGAIVLHDTGNFDCOSONERCDE 1212
Db 63 IENKIDITLNSKFKRKMSYSTGPDN---IDKNMSLIN----- 96
QY 1213 KSDGDNKRYAFRDPQDHDAACGCKSGSKRPTRYO-ITKKKAEKDEKTECVTDIILKEN 1271
Db 97 ---KHNEEMFNNNYQSFSTSSLIKONKTYVINAVAVSHILSLDR---INNGRNTS 149

QY 1272 DGKQVEDCHPKKNSNGYDPMOCGNINLVEDPVCAPPPRORIKLVAFLANDNEIKLOSO 1331
Db 150 SNNEVLSNCRNEKRG--MKWDCKKN-DRSNVYCIPDRIDLCIYNLS---IKITYK 201
QY 1332 VNLKEAFIKSAAEETFFSWYYSKSDGEGNEIDKEKIPAPLRSMFTYFGDYRDL 1391
Db 202 ETMKDHFIEASKKE---SOLLKKNNDKNYS-----KFCNDLKNSEFLDYGHIA 246
QY 1392 FGTDISKGHEGSKLKEQIDSLFKNGDOKSPNGKT---ROEWTSHSHIWEAMCALVK 1448
Db 247 MGNMDMFG-GYSTKAENKIQEYFKAHGSEISHKIKNFKEMWNEFRKIMFAML----- 300
QY 1449 ICAKKDDFTENGYNNVAFSDKSTLLEPAKRPQLWLTWYDYCYTROKYLDVOEK 1508
Db 301 -----SEKKNIN-NCKNIPQEL---QITOWKEWHGHEFLERDRSRKPSK 345
QY 1509 CKNNDOLK-CDTECNKCEDEYKY-MKKKEWIPQDKYKDRDKRRDRHIGVMTDY 1566
Db 346 CKNNTLYEACEKECDPCMKYRDWIIIRSEFEMHTLSKETETOKYKE----- 392
QY 1567 TGTNATDIYNKRFYASCGRPSASVQORNIQLEKQAYYDADKHCQ---TKPIE--- 1619
Db 393 ---NAENYL-IRISENKND-----AKVSLNNGDAEYSKYCDCKHTTLVKSVLN 439
QY 1620 -ND---DRTYNISSKDKCK-GLVKEA-NTGAIKQNNGNPNYNNKLTEDYLPSPRL 1672
Db 440 GNDNTIKEREHIIDDDSKFGCDKNSVDTNTKWECKNPTLS---TDVCPVPRQ 494
QY 1673 RICHFALDGNVTDPEVKDENGK--RKRLMEVAATEGYNLGQYKKEKKEKIKTSDAHKY 1730
Db 495 ELCLGNIDRIY-----DKNLLMKREHLALAIYESRLIKKYNKDKDE----- 538
QY 1731 SEVPPGSAKYSFYDLDIILGIDNLEDEKQTEENIKLTFKNKGSVSGSDSTGNP 1790
Db 539 ---VCKIINKTPADIDIDIGTDYMD---LSNRKLVGRKINTNSYVHRNKKN----- 585
QY 1791 GSTARKFPMNENKCEVYNNAMICGYRGRDNGNSNSARSDDLKKGCSVSDDYPRGKN 1850
Db 586 DKLFRDEWMAKVIKDDVNVVISWVK-----DKTYCKE 617
QY 1851 RDEGAYOFLWPAWEGDFCKHKELEKLVGACNDYTGCDNEKRRKCTDACTOYKRF 1910
Db 618 DDIENIPQFFRWFSEMGDDYQCDTKMIEFLKVCCKPEDD---NKSQNSYKEM 672
QY 1911 ISEWPOYEROKIKYGE---NKDKIYSEHPVAKDADAIEYLDKOLKICENKSGDCET 1966
Db 673 ISKKEEYNKQAOYQYQGNNTKMTSEKSIKPEVYLAKYSEK-----CSNINFEDEE 727
QY 1967 K-----C-MKDYSTORLTDGNSQNMPSLDEPEKV---EGKNCQOVPRG 2007
Db 728 KEELHSDYKKNKCTWCPEVKDVPIS-IRNNEQTSQEAIVPEENELNHRITPISSECPK 786
QY 2008 PPRVRETPSPRVLISKASKKEAKTAPPTQPKKVENLTTENRAQTTRAAQOTRK 2067
Db 787 MEQKERDDDS---ISKISVS-----PENSRRP-----ETDK 814
QY 2068 RTSTATTESDVG-TYVKAISNKPDSRGIEGNCNFTYGOYKWCIGYKSEKENGIC 2126
Db 815 DTSMLTKLKGVDVISMKPAVIGSSPN---DNIN-----VTBOGDNISGV- 855
QY 2127 MPPRRKLCINNIOYLYENLEKRDNDIEKAFIKCAIETOFMLKYLITENPAENELON 2186
Db 856 ---NSPLSDVVRPKKELEDONSSESEETVY-----NHISKSPINNCGDS 899
QY 2187 GTLPDEFKRIIMTYGYDKDMFGDISNDKLIITVNSVTLINNNKKNKDKKDEEL 2246
Db 900 GS-----GSATVSESSSNTGLSIDDRNGDITVRIQDTANEDVIRKKNAPKDEE 951
QY 2247 RKFWEKNNKFTWEGMIGLYTLHTDEN---EKEKIRNDYOYNDMTKLTPLSEFVVRK 2302
Db 952 KGADEER-----HSTSESLSPDEKMLTNGEGNSLN-----HEEVKEH 990

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:32:15 ; Search time 79.1271 Seconds
(without alignments)
3586.886 Million cell updates/sec

Title: US-10-087-013-2

Perfect score: 19407
Sequence: 1 MGFSCRYFLIKMGNAASLE.....IDLINDSLVLTLLIYMKY 3542

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

al number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications, AA.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	2972.5	15.3	2710	9	US-10-153-273-12
2	1032.5	5.3	700	9	US-10-153-273-10
3	665.5	3.4	1086	10	US-09-924-154-15
4	633	3.3	921	9	US-10-153-273-8
5	617	3.2	1435	9	US-10-153-273-4
6	599.5	3.1	1421	10	US-09-924-154-13
7	594.5	3.1	1501	10	US-09-924-154-17
8	591.5	3.0	1143	9	US-09-924-154-14
9	431	2.2	749	9	US-10-153-273-6
10	382.5	2.0	972	10	US-09-924-154-16
11	365	1.9	411	9	US-10-153-273-20
12	343.5	1.8	1115	9	US-10-153-273-2
13	291	1.5	6281	10	US-09-815-242-12996
14	289	1.5	1639	9	US-10-087-464-10
15	287.5	1.5	2125	10	US-09-919-172-29
16	279	1.4	362	9	US-10-153-273-18
17	279	1.4	2789	10	US-09-801-574-57
18	276.5	1.4	2285	10	US-09-933-183A-2
19	276	1.4	411	9	US-10-153-273-19

20	273.5	1.4	311	9	US-10-153-273-21	Sequence 21, Appl
21	256	1.3	1605	9	US-09-820-843A-73	Sequence 73, Appl
22	234	1.2	2478	10	US-09-815-242-5816	Sequence 5816, Ap
23	234	1.2	2478	10	US-09-815-242-12967	Sequence 12967, A
24	230.5	1.2	3158	10	US-09-815-242-12611	Sequence 12611, A
25	225.5	1.2	2139	9	US-10-023-219-4	Sequence 4, Appl1
26	225.5	1.2	2139	10	US-09-727-384-6	Sequence 6, Appl1
27	218.5	1.1	2785	10	US-09-801-573-8	Sequence 17, Appl
28	218.5	1.1	324	9	US-10-153-273-17	Sequence 50, Appl
29	215	1.1	1331	9	US-10-087-464-50	Sequence 5703, Ap
30	215	1.1	2025	10	US-09-815-242-5703	Sequence 108, App
31	214	1.1	807	9	US-09-820-843A-108	Sequence 5635, Ap
32	213	1.1	2368	10	US-09-815-242-5535	Sequence 12389, A
33	213	1.1	2368	10	US-09-815-242-12389	Sequence 7, Appl1
34	206	1.1	1400	10	US-09-764-176-7	Sequence 1, Appl1
35	206	1.1	1404	10	US-09-811-045A-1	Sequence 493, App
36	201.5	1.0	1095	9	US-10-076-622-493	Sequence 493, App
37	201.5	1.0	1095	12	US-10-007-805-493	Sequence 565, App
38	200	1.0	1341	9	US-10-076-622-565	Sequence 573, App
39	200	1.0	1341	12	US-10-007-805-565	Sequence 573, App
40	200	1.0	1349	9	US-10-076-622-573	Sequence 5834, Ap
41	200	1.0	1349	12	US-10-007-805-573	Sequence 553, App
42	199	1.0	2437	9	US-09-815-242-5834	Sequence 553, App
43	198.5	1.0	1013	9	US-10-076-622-553	Sequence 475, App
44	198.5	1.0	1013	12	US-10-007-805-553	
45	197.5	1.0	1002	9	US-10-076-622-475	

ALIGNMENTS

RESULT 1
US-10-153-273-12
Sequence 12, Application US/10153273
Patent No. US20020169305A1

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chilins, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:
ADDRESSER: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH21.1FM0V1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium falciparum
 SEQUENCE DESCRIPTION: SEQ ID NO: 12:
 US-10-153-273-12

Query Match 15.3%; Score 2972.5; DB 9; Length 2710;
 Best Local Similarity 26.5%; Pred. No. 2,6e-155;
 Matches 865; Conservative 442; Mismatches 873; Indels 1083; Gaps 120;

QY 9 IIRGNASSLEEDAKSPIIKESHKSANVLEFVAKNIRHPSK-YAKEHVDLTKGDTJTK 67
 Db 6 LAKGPEALAGDIED-----ESAKHMFDIGVDYKVEEKEKRGKGLQRLSEA 58
 QY 68 EFRGSPTPYNNKINYYPPPCNDIHEHTNLRDYDNLHPCHGRRQNRDEDESEEC-G 126
 Db 59 KFEKNESDPQTPPD-----PCDIDHRYHTNV-----TTNVINPCADRSVRSDEYGGQCTH 110
 QY 127 NKTINRYKKNDAIACAPRRHRHMCNLEALNDINTONHIDLGNLVTAKEGSEIVNN 186
 Db 111 NRIKDSQGGNKGACAPRYRLHVCODNLEIEPIKTITNHLVYDCAKKEGOSITOD 170
 QY 187 HP-HKGT-----SDACTALARSFADIGDIYRGIDMFKPVNH-----KYETGLREVFKE 235
 Db 171 YPKQATYSGSPSOICTMLARSPADIGDIYRGHDIYLGHPQEIKNQQLKTTIFGI 230
 QY 236 HD-----GMEDEVKNDYNDPSGNYKLEBAMNANRKKVKEALTCDAISYSGTFMOSEST 292
 Db 231 YELYNLAEARLYGND--PE-----FFKLREDMWATNRETVAIKATCA-NWNTYF-HATCMR 282
 QY 293 PLESNKCGHKGKVPYTNLDYVPOYLRFPEDEWEECECRKRNITKLKKVKDCR-----NDKE 348
 Db 283 GERTKCYCNCNDQVPTDYDYPOYLRFPEDEWEECECRKRNITKLKKVKDCR-----NDKE 342
 QY 349 RLYCSHNGHDCITTTWKKGIILHDKCSDCTCKCYFEVWILGNOQEPKREKETEKO 408
 Db 343 R-YCSHNGYDCETKRAIKRLYRGKOCISCLACNPYVWMINOKEOEPKREKETEKO 401
 QY 409 SY-----LSNDNKFVNINSE-YKQFYEKLENOYATNDFEFLNLNGKVC-----454
 Db 402 KYENGASGSRQKRDAGTGTITNYGSEKFTDELANKSTRYVDLFLEKLSHEELCTYVK 461
 QY 455 --KGLPEKIDITFTNSA-----DNGIIFYRSEYCOVPCDGYKCDGIKTYH-KSPND 504
 Db 463 DEEGTIDIFKNVNSDSTSGASGTNYESGTFYRSKYCOPCPYGVK-----KVNNGSSNE 517
 QY 505 RERVNN-----EDKPPMGVAPRNTIYLYSGNEGODITOKLENCN-----SS 547
 Db 518 WEKNGNGCKSKGLYEPKDEGTTITLKSGHDIIEELKNKPCDERNGDTINSGG 577
 QY 548 TNYKDKNN-----OKWECYKDEINRCKLEONTEINNBPK-----II-----586
 Db 578 TGGSGGSGNSGRLEEMKC-YKGEDVYKVGDEDEDEYENKAGGLCTILKOKKAKE 636
 QY 587 -----SHNFEWLYTLRDTIKWDLKTCI-NNTTTHC-IDECNHC 629
 Db 637 EGGNTSEKEPELQKTFNPFYFVVAHMLKDSIHKKKLOGLCLONGNRIKGNKKNCCND 696
 QY 630 LCFDRVWOKKEEEMSIKLFTEKKKNIO-----QSYSSINNLFEGYF---672
 Db 697 ECFKRWITQKDEMGKIVQHP-KTONIKRGSGSDNTAELLIPFDHLYQVNLQSEFLKGD 755
 QY 673 -----FVYMDKLDKDEAKWELENIKRKKNEFSNLENND-----YLENAIELL 718
 Db 756 SEDASEEKSNSLDAEER-----BELKHLREIIESENNOBASVGGVYTQKIMMDLL 809
 QY 719 DHEKETAIC-----KDNNTNEACETSHNATTPCVRKRGGTQ-PTKNIKETIQ 766

Db 810 NYEKDEADLCLEIHEDEEEKEKGGNECEIEGENFRINPCSGSEGNKRYPYLANKVAYQ 869
 QY 767 YFKRSAYEAPRNRGLHKLKGAHEGITYKRGRRKDFKDMICRIMIKHSNRNLGFSNGPCD 826
 Db 870 MHHKAKTOLASRAGSALRGDISLAQFKNGRNGSTLKGICIKINENYSNDSNGSGCCT 929
 QY 827 GKGTGGDIOTRFVYGTGEMVEVDEEHNKKHEDYIMPPRRHITCSULEHLQTDHPLNGNI 886
 Db 930 GK-DGDHGVARRHIGTEMENISGKQTSYKNVFLPFRREHCTSNLENLDV-----GSV 982
 QY 887 V-DLIVNNSFLDVLISAKYEANKIIRMYKEKNKNGKPEVNDP---KHQTTICARAYS 942
 Db 983 TKNDKASHSLDGYOAAKTAADAEIIRKKQDNNI-----QLTDPOQKQDEMKCAVAYS 1038
 QY 943 FADIDIGIRGLMERNGDMVLOGLHLETVGNHKSILKG-KGNDKYND--APRYLKL 998
 Db 1039 FADIDIGIRGRMDWEDSSTDMETRLITVFNKIEKHGDIKDNKRYGDESKKPAAYKL 1098
 QY 999 RENWNEANRAKYWEAMKCDIKYLKDSGHOSHOSSYSGYSDHTPTDDYIPQKLRMTEMA 1058
 Db 1099 RADWWEANRHOVWAMKC-----ATKGLIC--PQMPVDDYIPQRLRMTEMA 1143
 QY 1059 EYCYVQKKEVDKLEKECKECKDNQGGCTYESGTGCTKTEACNEYNDIITGLKREQN 1118
 Db 1144 EYCYCAQSQEYDKLTKICADCKSKGDG-CT-QGDVDCGKCAKADKYKEIEKMEQMR 1201
 QY 1119 IISDKYELHQAQSVNSGIEASTAKNHID--RNVIEFLSELVQONGK---SNK 1171
 Db 1202 KISDYNMLYLOAKTSTNPG---RTVLGDDPDYQDMVFLPIPHKASIAAVLYKRA 1257
 QY 1172 SGTDSASAVIGNTTYENVAGYVILHDTGNFDDCQSOSENEFCDEK-----SDCKDNKAYFR 1225
 Db 1258 AGSPTEIAAARPIPYRAAGYIHQELIGYGCGQDQCEKHHATISTITKEKKEFTFK 1317
 QY 1226 DKPOHDACGCKSKSKSTRVQITKKAABEKTCTVNDIILENDCKQVDECHPRKN 1285
 Db 1318 QPPEYATACCKINRSQTE---EPKKEEYVESACKIVEKILLEGKNGRTVECNPKES 1373
 QY 1286 SNGYEDMOC-GNINLVEBPVRCMPBRQKLCVHLADNENKIKLOSOVNLKAEATKSA 1344
 Db 1374 ---YPMDCKNINIDISH-GACMPRRQKLCLYIAHSQETENIKTDNLKDAIKTAA 1429
 QY 1345 ETFPSWYTYKSK-DGEGNELDKLEKGIKIPAPLRSNFFYFGDYRDLFFGDISKGE 1403
 Db 1430 ETFPSWYTYKSKNSSEAKILDR---GLIPSOFLRSMTYFGDYRIDCLMTDISKKN 1485
 QY 1404 SKLKEOISLFEKNDDOKSPNGKTRQEWTEHSHIWEAMCALVYKICAKKD--FTENT 1460
 Db 1486 AKADKIGKFFSKDGSKSPSGLSHQEWKTINGPEIMWGMICALTKYVTDIDNKRKIKNDY 1545
 QY 1461 GYNNVKEFSK-STTLEEFARKPQFLRMLEWYDYCYTRQYKLDVQEKKS-NDQLKCD 1518
 Db 1546 SYDKVYNSONGNPSLEFPAKPOFLRMIMWGEFCAERKKEKIKIYDACEINLSNQCN 1605
 QY 1519 T---ECKKCEDYKTYK-KKEWIPQDKYTKDERDKRPRDROHIGVMTDYGTATDY 1574
 Db 1606 DAKHRCQAQRAQOYEYENKKEFSGQTNFVLKANVQODPEYKGYEKD-----1656
 QY 1575 LNRFTASGDKRGSASVYVQNRIOLEKQAYYDADHCCTKTLENDKTYNISCKDK 1634
 Db 1657 -----GVQP-----LOGNEYLLQK--CDNNK-CSC-----1678
 QY 1635 GLVEANTGAIKMONKPNYNNILKELTEDVLPSPRLRICFHALGNTYDPEVKENG 1694
 Db 1679 -----MGNNVLSPKPK---1691
 QY 1695 KRLMEVATEGYNLGOYTYKEKKEKIKITSAHKYSTEYVPCSAANKYSFYDLRDIITIG 1754
 Db 1692 -----PFGKY-----AHKYP-----1701
 QY 1755 DNLDEKQKTEENMLKIFNKNGTSVGKGSSTJGNGSTARKPFWMENKECYWNAIMIGY 1814
 Db 1702 -----1701

QY 1815 KGRDDGNSGNSARSDLDLKKCGVPSDDYPMGNRDEGTAYOFLRMFAWGEDECKRH 1874
 DB 1702 ----- 1701
 QY 1875 EKELELVGACNDYTCGDNEDEKRRKCTDCTOYKKEISEMKPOYEKOIKYGENKDIYS 1934
 DB 1702 ----- 1704
 QY 1935 EHPVADAEADAREYLDKOLKICENKSGDCEYKCMKDVSTQRLDGNSONMPASLDDEPK 1994
 DB 1705 ----- 1706
 QY 1995 EYEGKNCVPRGPPRVRETPSPVSLISKATASKKEAKTAPPTOPKVENLTTEMRA 2054
 DB 1707 -YOGK---HYPSIIP----- 1725
 QY 2055 QFTFRBAAOQTRRTSTATTESDVTGTMVAKILSNKPDSCNGIEGONPKTYGOY--PKWG 2112
 DB 1726 ----- 1766
 QY 2113 CI-----VGSKENENGICMPRRKRLCINNIOYLANET-----ENKRD 2151
 DB 1767 CIPSDTKSAGATGKSGSDSGSICIPRRRLIYVGKLO--EWATAPQGEAAPSISRA 1824
 QY 2152 NDIKEAFICAIETQFLMLKTIIE--NPAENE-----LONGT 2188
 DB 1825 DDLRNAFIOASAIETFLMDRYKEKKPODGSQALISOLTSTYSDEDEPPDKLONGR 1884
 QY 2189 IPDEFRIYTYGDYKDEFEFGDISNDKIIIVTNSVTTLNENKKKODKDELRK 2248
 DB 1885 IPDFELRMETLGDYRDILVHGNTSDGNTGNSNNNLVLEASGNKEMOKIOETQ 1944
 QY 2249 I-----FEKNNKFIWEGMIYGLATY-----HLTDNEKEKIR 2280
 DB 1945 ILPKNGTFLVPSAQTPTPKMWNHAEISWKMICALITYTEKNPDTSAAGDEKNLEKD 2004
 QY 2281 DNT-----QYN-DMTKL-----TPSLEEFYKRP 2302
 DB 2005 EYVEKEFGSTADKHGTASPTGTGYTOYDEKYLKEDTSGAKTPSASDPLLDVFLRP 2064
 QY 2303 QFLRMTEMAEEFCNRKBDLKLFLAGCKEY-----CN---GSN 2339
 DB 2065 PYRTIEEMGQNFCKRKHLKQIKHECKVEENGSGSRGGITROYSGDEACNEMLPKN 2124
 QY 2340 DGKTQE-----CAECVYQNFILKWKTEYEROREKFKDKD----- 2376
 DB 2125 DGTVPDLKESCAKPCSYSYKWKIESKGEKEQKAYEQOKDKCVGNSNKHDFCETLT 2184
 QY 2377 -GKKYDY-----PSTERDEKATCAHEYILNMLKELCGKDCSCMOKPSQLPKTTQOS 2430
 DB 2185 TTSKADDFLKTLPCKPNNVEGKTIFFDDDKTFK-----HTKDDPCLKFEFVNCKK----- 2234
 QY 2431 QSSDAMDPESLDYVEEFKCEPSELKSKSGMIHTKRIEPIPMNCVKAAYILSKEA 2490
 DB 2235 -----DECD-----NSKGTCKRNK-----NSID-----ATDI 2256
 QY 2491 ENNMIDTLKEKFIPIESTKEKESKNSMTNNPCDPKPYADPYIGRRNCPEN----- 2543
 DB 2257 ENGVDSTY-----LEMRYSADSKSGFNGD-----GLENNCRAGIFEI 2295
 QY 2544 REENRRKVDYEMKCYKNSKFEYQEKRRVCYPPRRHMLCLNLDKIERLKSNTYLLKMR 2603
 DB 2296 RKD-----EMKC-RNVGCVY-----VCKP----- 2313
 QY 2604 RTARNEGIDILKMFNSENGCANMPIDTMYKSFADLGDIVRGDMLRIGGYLPVEIKLY 2663
 DB 2314 -----ENVNG-----AKGNHIIQIRALV-----RWY 2336
 QY 2664 KVEEYLYGKRNKNGKRNKYND-----VQTFRSAMDMANRDKIAMTCKAPEDAKFERKG 2719
 DB 2337 EYFEEDYNKIKHKISHRIKNGEISPCIKNCVEKMWQDKRE--WKELTERKD----- 2387

QY 2720 RMDGERITLIDQCGKHDDPVVDYIIPORFRMTENSEYECALMELEKFKSCDHCK 2779
 DB 2388 -----QYKNDNSDDDNVRSFLETLIPQITDAN---AKNKYIKLSKFGNSCG-CS 2432
 QY 2780 TSDRCKNDVDENKCEQCKTKPCQCKYKPNVLKWKLFIDQSKYKVELBQPIYTWISYIDH 2839
 DB 2433 AS---ANEQKN-----GEYKDAI-----DCMLKRLKD-----KIGCECK 2465
 QY 2840 QNFQKRLKTFKSCSVSEFSEYELHETSKCLNYKFENEDGSSNIRTYAFETPKSYKE--- 2896
 DB 2466 HH-----QTSDFECSDFIPQOTILED--ETLDDDIETFEAKKNMMPKICEVYLKTAQODE 2518
 QY 2897 -ACSTLPKSNPLDNCPTQONKD 2918
 DB 2519 GGC---VPAENSEEPATDSGKE 2538

RESULT 2

US-10-153-273-10
 ; Sequence 10, Application US/10153273
 ; Patent No. US20020169305A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sim, Kim L.
 ; Chitnis, Chetan
 ; Miller, Louis H.
 ; Peterson, David S.
 ; Su, Xin-zhaun
 ; Wellens, Thomas E.
 ; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 ; NUMBER OF SEQUENCES: 37
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbe Martens Olson & Bear
 ; STREET: 620 Newport Center Drive 16th Floor
 ; CITY: Newport Beach
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/153,273
 ; FILING DATE: 21-May-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/210,288
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fuller, Michael
 ; REGISTRATION NUMBER: 36,516
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 235-8550
 ; TELEFAX: (619) 235-0176
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 700 amino acids
 ; TYPE: amino acid
 ; STANDARDS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Plasmodium falciparum
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
 US-10-153-273-10
 Query Match 5.3%; Score 1032.5; DB 9; Length 700;
 Best Local Similarity 34.8%; Pred. No. 4.6e-49;
 Matches 247; Conservative 110; Mismatches 250; Indels 103; Gaps 22;

QY 136 NDALACPPRRRHKCDKNLEALNDINTONIHDLGNLVTAKEYGESIVNNH--PHKGT- 192
 Db 5 NKVACAPRYRLHLCYDNLESID--TSTTHKLLLEVCAMAKEGNSI-NHTHTQHRTN 61
 QY 193 ----SDACTLARSFADIGIVGIDMF-----KPVNDKVTGLREVFVKIHDGMEDE 242
 Db 62 EDSASQCLTYLANSFADIGIVGIDIGVGRKDLGYDNKEKORRKLQKDIKFIKTH--KDY 118
 QY 243 VKND-----YNDPG-SENYKRLREAMNNVNRNRYWEAITGDASYKSGYFMOSSNTPLES 296
 Db 119 MKTNGAQERVIDAKGDEFFQLEDMWTNKRREYWKALICHAKREANTFETKICANVGKT 178
 QY 297 NPKCGHOGVPTNLDVPOYLKRWDEWGESEFCRKRNIKLYKVDSCRNDKERLYCSHNG 356
 Db 179 NGCCHGCGVPTFYDVPQYLKRWFEWMAEDFCRKKKLENILOKCRDVEQLYCSGNG 238
 QY 357 HDCTTTMKKGIHLNDKCDGCTCKGVFEWVGNQOEAFKKRKEKEKEI----- 407
 Db 239 YDCTKTITKKGKLVIGEHCTGCSVWCRRMETWIDNQKKEFLQKRRYETEISGGSGKSP 298
 QY 408 -----OSYLSNDNKFVNNINSSEYKQFYKLETOYATNDTFLNLNEGKYGCG-- 456
 Db 299 KRTKRAARSSSSDDN-----GYESKFYKRLKEVGYODVDKFLILNKEGICQKQ 350
 QY 457 -GLPGEKDIETMNSADCKGIFYSEYCOYCPGCVACDGIKTYHKSDNDRERYNNEDYRP 515
 Db 351 VNGEKADNVDFTEKTYK-TFSRTEICEPCWGLEKGGPPMKVKGDKTCSGSAKTYTYP 409
 QY 516 PMGVKPTNTVLYSGNEQDITOKLEFNCSSSTNYKDNKNQWECYKQD-----ENINCG 570
 Db 410 K--NITDIPVLYPKSQNLKRYKNFCEKGP--GGGQIKKQCYDDEHRSSKNNNC 465
 QY 571 KLEQNEINNDNPKIISNFELWVYTYLLRTIKMNDKLCYCINNNTT---HCIDECN 626
 Db 466 VESTMDKETOQKQVYSYVFEWDMVHMLHDSVEKTELKSCINNNTNGTCRNNKCK 525
 QY 627 RNCICFDRWVKOEKEEENSINKLETKKKNIOQSYSNINNLGEYFFKYMOLDLDEAKM 686
 Db 526 TJCGCGQKVEKQOEMAKIDHFGQDIYQO-----KGLVSPDYGLDLY---LKG 576
 QY 687 KELMENIKKKKNEFSNLENNRDYLE-----NAIELLDHLKETATICKDN 732
 Db 577 GNILONIKDVHGDIDIKIKILDEDAVAVVGLGKNDITDKILOHKEBAEQCKQK- 635
 QY 733 TWEABE-----TSHNATNCPVAPRGSTOPTKNIKELAOYFKRSAYEE 775
 Db 636 -OECEKKAQOESRGRSARETREDERTQ--OPADSAGEVEEEDDDDYDE 681

DLT 3
 US-09-924-154-15
 : Sequence 15, Application US/09924154
 : Patent No. US20020127241A1
 : GENERAL INFORMATION:
 : APPLICANT: Natum, David L.
 : APPLICANT: Sim, Kim L.
 : TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
 : FILE REFERENCE: 05213-0465 43170-262105
 : CURRENT APPLICATION NUMBER: US/09/924,154
 : PRIOR FILING DATE: 2001-08-07
 : PRIOR APPLICATION NUMBER: US 60/223,525
 : NUMBER OF SEQ ID NOS: 17
 : SOFTWARE: PatentIn version 3.1
 : SEQ ID NO 15
 : LENGTH: 1086
 : TYPE: PRT
 : ORGANISM: Mammalian
 : US-09-924-154-15

Query Match 3.48; Score 665.5; DB 10; Length 1086;
 Best Local Similarity 21.08; Pred. No. 1,5e-28;

Matches 312; Conservative 186; Mismatches 431; Indels 555; Gaps 60;
 QY 2070 SRTATTESVGTWVKALISNKPDSRGLEGCPNPKTYGQYPMGCIYVGSKENENGICMP 2129
 Db 7 SKSIIRNKDVDSIKDPEKNKFIDYSLYECGKTI--KEMKICIDNQFK--SNLCAPI 62
 QY 2130 RRRKICINNOYLANETEN-----KRD--NDIKFAFKCAIETQFLMKTYIENPAEN 2182
 Db 63 RRIQCLTVNI-ILFSENEEYIYKNDISNNKRENIILAVKLESMLYOKH-----N 113
 QY 2183 ELONGTIDPEFRIMYYGYDKDMFEGTDISNDKKITVTVNSVTTLNE--NNKKQDK 2240
 Db 114 NEYNSKLCDDIR-----WSLVDYGDIIIGRDLYKNNTDYIKQFQKIRNNEYNNELNDE 169
 QY 2241 -----KDEBLRIFWEKNAKFLWEGMIGLYLHYLHDENEKERINDYONDTKLTP 2293
 Db 170 LNNELNDEKNIKLRKEMWEKEDIMEE-----TKHND--KRIE 208
 QY 2294 SLEEFVK-RPOFLRFTWMAEEFCNRKREOLIKLEAGKEYECNGSNDGKTOGEACVYT 2352
 Db 209 KCKYFAKDEPOLVRIEERSKQFLDEKNYMLFTLNTYENMIHENN-----CKQ 259
 QY 2353 YONETKKMTEYEROREKFKDKDGK-----YKDPSTERDIEKATCAHEYLMKL 2404
 Db 260 YNKWYQNRKKEWTELSNEFNKIFPERNYQIHSNIFKEYKKNVNDIIGTLNTEYNNF-- 317
 QY 2405 KELCGNKCSGKORSSQLPPTTQOSQSSDANDPESLDYVPEEPNKECELSKSGMT 2464
 Db 318 -----CKEKP-----ELVSAK----- 329
 QY 2465 HTKITEPRIPANCVEKAAVYLSKEAENNDITLKEKPIESTREKESKSNWTNNPCD 2524
 Db 330 YNLKAPNAKSP-----RIYKSKHESSVFGCKTI----- 360
 QY 2525 PKKPAPPKTYGRNPNCEENRENFKYDEYKCYKNSKFTYQEKRVCPVPRHEMLRL 2584
 Db 361 -----SKYKKKWNCKSNK--YTKPBGVGGPPRROOLCGIYI 395
 QY 2585 DEIK---IERLSDSNYLLKWRRTARNBGIDILNFENSEGCAANP-ICDMKKSFAIDG 2641
 Db 396 FLINDGNEGLKDH-----INKAANTFAHLKAKYENAG--DKTCMILISYADIGD 446
 QY 2642 IVRGTDMLRIGGLYPVEI---KLYKVEYTYGWRNKNRKNKYNDYQFERSAMANDR 2698
 Db 447 IVRGIDVWR-----DINTNKLSEKQKLIPOGGSRRKQNDNNE--RKNWMEKQ 494
 QY 2699 KDINKAM-----TKAPBDAKLFRKGRMDGFERITLLIQDKCGHKDDPPVDYIPQRF 2750
 Db 495 NLIWSSWVKHLPKGTCK-----RHNNFK-----IPOFL 524
 QY 2751 RMTWSEYCYCALMELEKFKKSCDHCKTSDCKNDYDENKCEOCRCQGEYKKNVLMW 2810
 Db 525 RMLKEMGDECEMGTEVQLEKICENKNSER-----KCKNCSSTIEKAIKER 573
 QY 2811 KSLFDIOSNKYKELYEQPIYTKISTYDVONVQOKLTERKSECSVSFSEYHETSK-CL 2869
 Db 574 KNEYNLOSCKFDS--DKLKNKNNLYN-----KFPDSKAYLRESKQGS 615
 QY 2870 NYKFNENDGSNIRTYAFEEYTPKSYKAGS-CTLPSKN-----PLDN--OPTQDNKGCCE 2922
 Db 616 NIEEND-----ETFLF--PKYKACAVGCENPSSKALFKTYNVPFIESKR--SE 663
 QY 2923 LQTFPCSKNDYNNLDNMMNAYLYVNSSDDNKGVLIPRRRH/CTPRTIAYNVRKDKET 2982
 Db 664 LSSLITDKSKN-----TPNSSGGG-----NY--GDRQI 688
 QY 2983 LKK-----KLTSAFSGQLLGQKYSEELC-----FEAMKSYADYDIKGTDM 3029
 Db 689 SKRDVYHHGPKKEVKSGEKVEPRIDAAVTENEFTSNRNDIBKREKSKDHSPPVASKDI 748
 QY 3030 MDTSLSEKIKLIFETSEATENKRTWMENNRRQIWMHMLCGYIANSKYTLDBGMOQLPK 3089
 Db 749 -----KNEEPQRYVSEN-----LPK 763

```

QY 3090 DEETNPLRLIMAKOACKKRVSDLTCKCPRSNEDNFESALLROPCONDIRKTY 3149
DB 764 IE-----EKMESDSIPI-----THIEA-----EENVOYIK 782
QY 3150 SLNLIKNTMENTNIRYKOLKDOSSGNIDNKP-----EENVOYIK 3191
DB 783 -----KGOSSNSDNDP/VVSGRESKDVNLTSEIKENEKGVIK 822
QY 3192 SKDSOCALF-----NDINEIVTGTKNENNE-----FKEVLLKLYPGLYVED 3235
DB 823 TDSSKSIETSKIPSDONNHSDLSQANEDSNQNKETINPSTENKLIKIH----- 874
QY 3236 ETKKNHVLGNITKEEQIYRPAKLYFTTPVDSFYQAPLESTRVAVQYDPKNDILKSS-- 3293
DB 875 --KTSDSDDHGSKIKSEIEPEKLEESPLTK-----KTESAIGDKNHESVKSADI 925
QY 3294 -----ISVIVY-----SALGLIALHFMKKRKSVDLRLININQOGEVGP 3334
DB 926 FQSEIHSNDNRIRIVSESVYVODSSGSMSTESTIRTDNKKDKTSEDIAPISTIN----- 976
QY 3335 TLESKNRYIPRSGPYKTYIEMGDTSGDEDKYMWDLSSDITSESEYEELDINDIY 3394
DB 977 -----GHEKIGSSADDRSEDEKSIIDKSENFENKSSHSDIKQSD-- 1017
QY 3395 VPSGPKYKILIEVYLPESKRDIPS-----DDTSPNDTPRTN 3430
DB 1018 NEGSTDYESLTE---ESPKGDLSEVSPSSIDMDLKRNKSSPVTS 1058

```

RESULT 4
US-10-153-273-8

Sequence 8, Application US/10153273
Patent No. US20020169305A1

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.

Chitnals, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US

ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH21.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids

```

? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? HYPOTHETICAL: NO
? ORIGINAL SOURCE:
? ORGANISM: Plasmodium falciparum
? SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-153-273-8

```

Query Match 3.3%; Score 633; DB 9; Length 921;
Best Local Similarity 27.8%; Pred. No. 7.3e-27;
Matches 224; Conservative 113; Mismatches 264; Indels 206; Gaps 42;

```

QY 948 DIIRGDLWRNDGMYLQGLHETVFGNIHRSKRGKNDKYN-----DDAPKYLKLENNY 1002
DB 80 DYITASDI-ENGNSI-----GNIDMYSVDKDNANFNGLDAGSANIRKIRKEQ 128
QY 1003 WEANRAKVMAMKCDIKYLKDKSGHOSYSSYGYSDHPPLDITYPOKLRMTWMAEWYC 1062
DB 129 WKC-AKV-----CGLDVCGLKNGN-----GSIDKQKQIILIRAL--LKRWEYFL 171
QY 1063 KYQKKEVDKLEKCKEKKDKNQGGCTKSGTGCTGTCTACNEYNIDIGLMEQWNTSD 1122
DB 172 -----EDINKINAKISHCTKKNESSTCTNDCPNKCT-CVE-----EWINKRTWKNIKK 220
QY 1123 KYRELHE-----QAQMSYSN-----SGIEA-----SSTAKNH 1149
DB 221 HYKTONENGDNMNNKSLVTDLILGALQPSDVNKAIRKPSGLTAFESFGGLGANGNSEKKEG 280
QY 1150 ID-----RNVIEFISELYOONGKSNKSGTSDSAVIGTITVENGCAYLHDTGNFD 1201
DB 281 EDDYDLVLCMLKNEKQIOEKKKRGHETSVENG--GKCTPLDITLTLTE-----E 327
QY 1202 DCOSONE-----FCDEKS--DGKDNK-----YARDRKPDHGDAGCGKSGSKPTRYQ 1247
DB 328 PIBEENOVENAPNICPQTVEDKKEEETCTPASPPEKPVH--VARRITTPPEVEK 365
QY 1248 IKTAKRAEERDTECKTVNDILKENDGRQYEDCHPKKNSNGYDPDQGNINL-VEEDPRVC 1306
DB 386 I---WRGRNKRKTCCEIYVAEMLIKDKNGRTYGEVYRKET---YSEWTCDESKIKMGQIGAC 439
QY 1307 MPRRQKLCVHFLANDNEIKKLOSQVN-LKAPFKSAAAETPFSWY-YYSKSGEGNELD 1364
DB 440 IPRRQKLCVHFL--EKIMTNTNELKYAFKCAAAETFLIMONYKDKNGNEMEDLD 493
QY 1365 KELKEGKIPPAFLRSMFYTGDRDFFLGTDI-----SKHGEGSKLKEQIDSLFKNGD 1418
DB 494 EKLKGIILPEDEFKROMFYTRADYRIDCLGTDISSKDTSGVGC---KVKCNIDVF--- 546
QY 1419 QKSPNG-KTROEWWTESHEIWEAMLCAL-----VKIGARKDDETFENYGYNNVAF-S 1468
DB 547 YKISNSIRYKRSWENNGPIVMEGMLCALSYDSINNVPNETIKKLTGEGNNNEKYIFGS 606
QY 1469 DKSTTLEERAKRPQFLMLTEWTDYCYTRQKYLKDOEKCKSNDOIKDCTECNKK--C 1525
DB 607 DSSYTLSEFSEERQFLMLTEWENEFCEQKKEKYLALCKPCD-VDGGKCKNGKCVAC 665
QY 1526 EDVVK-----YMKKKKEMIPQDK--YYKDRDKRRRD--RQHGVAVWTDVDTGT 1569
DB 666 KDQCKQYHSWIGIWDNYKKQKGYTEVKKIPLYKEDKDVNSDADADYLTQDQNNKKV 725
QY 1570 NATDYLN-----RKFTASCGDRPGSASVYQRIQLLEKQAYYADADHCGCTFIENDDK 1623
DB 726 NGTDENCEYKCKMHTSSTSDMPES-----LDEKPEYKDK-CNC---VPRECN 771
QY 1624 YTNISS-----KDKCKGL 1636
DB 772 ALSVSGSGFPDGOAFGGGVLEGTCKGL 798

```

RESULT 5
US-10-153-273-4

[illegible]

Db	202	ETWKOHFIEASKE	---	SQLLKKNNDKNNYNS	-----	KFCNDLKNSLFDYGHIA	246		
QY	1392	FGDIDSKGGBESKLEQIDSL	EFKNGDOKSPNGT	---	ROEMWTESHEITWAMLCALV	1441			
Db	247	MGADMDG	-GYSTAKENNIOEFKAGHSEI	SEHKIKNFKEWNEEFREKLEAML	-----	300			
QY	1449	IGAKKDPEFENGYNNVFE	SDKSTLEJESFAKRPQELM	EWMDODCYTQKTLKVOEK	1500				
Db	301	-----	SEHKNNIN	-NCKNIPQEL	-----	QITOMIKEMGEFLLEDNRNSKLPKSR	345		
QY	1509	CXSNQOLK	-CDTECKKCKEDYKY	-MKKKWIPDQDYVYDEHDKRFRQHIGVATDY	1566				
Db	346	CKNNLTLYEACEKECJ	CDPCMKRYDWIIRSEKFMHLS	KEYETOKVPE	-----	392			
QY	1567	TGNNADTYLNRFETASG	DKRGKSGASVYQNR	QJLLEKAYYDADKHGCC	---TKFIE---	1611			
Db	393	-----	MAENLU	-IKISNNKD	-----	AKSLLLNCAEAYSRYCDCKHTTLYSVLN	439		
QY	1620	-ND-	---DKYTNISSKDC	K	-GLVKEA	-NTGALIKOMKRGNNNTNKL	ETEDVLFPSRUL	1672	
Db	440	GNDNLTKE	REHILDDPSKFGCDKNS	VDNTKWECKNYILS	-----	TKDYCVPRRQ	494		
QY	1673	RICFHADGNTDPEV	KREBNLU	---RKRLEVAAT	EGYNLQYIKERKEKIKITSDAHY	1730			
Db	495	ELCLDNDRITY	---	DKNLMTKEHILAI	AIYESRLIKRKYKNDDKE	-----	538		
QY	1731	SYEVPSCAMKSYFD	LIDILIDNLEDEKOTEENL	KIFPKNGSTVGKSDSTGTNP	1790				
Db	539	-----	VCKIINKTFAD	IIDJIGGDIYND	---LSNRKLVG	INTNSKYIYHRNKN	-----	585	
QY	1791	GSTARKEFWNNEK	ECVWNAIMCYKGRGRD	SGNSARSDEDLKCGSV	SPDDVPMGN	1856			
Db	586	DKLFDEDMWKVY	IKDYNVAVISWPK	-----	DKTYCKE	617			
QY	1851	RDEGAYOFLMFA	EMGEDECEKHEKELEK	VACANDYTGDNEDKRRKCTACQYKFF	1910				
Db	618	DIENITPOFFH	PESEMGDDYQODKT	KAIETLAKCEKCEPEDD	-----	NCKSKCSYKFW	672		
QY	1911	ISEMFPYOKEOIKKYE	---	NNDKIYSEHPAKDA	EADAREYLDKOLKIKENSGDCEY	1966			
Db	673	ISKKEEYNKOAKOY	EOYOKNNYKMYSEK	SIKPEVLYLTKYSEK	-----	CSNLNFEDF	727		
QY	1967	K	-----	C	-MKVYSQRLD	SGNMMRPSLSDDEPREV	---EGKCNCQVPRG	2007	
Db	728	KEELSHDYKNC	CTMCPEVYKDPIS	-IIRNEQOSQ	BAVEEENETLAHRT	EPISISEPRG	786		
QY	2008	PPVARETPE	PSRVSLSKATASKEAK	EAPTQPKVEVNTTEMBAQT	RFRBRAAQPRK	2067			
Db	787	NEOKERDDSD	---	LSKISVS	-----	PENSRR	-----	ETDAK	814
QY	2068	RFSATITTESYV	-TMYKALLSNK	PKPSRGISBCNPKTYGQ	PKMGCIYVSGSKENENGIC	2126			
Db	815	DTSMLLKKG	VDISMRAVIGSSPN	---DWIN	-----	VTBEGDNISGV	855		
QY	2127	MPPRKRLC	INQIOLAYEFENKRDNDIEA	FKCAAITQFLMKYI	INBPAENLQIN	2186			
Db	856	-----	NSNRPLSD	VPAPDKKLELDON	SESEYIV	-----	NHLSKSPSINNIGDS	899	
QY	2187	GTIDDEFERRIM	YTYGDYKDMFGT	DISDKKITVYNSVTTIL	NENKKNKODKKDEEL	2246			
Db	900	GS	-----	GSATVSESSS	NTGSLSIDDRNGD	PFVFTODANT	EDVIYIRENADKDEDE	951	
QY	2247	RKIFEMKKNK	FIWEGMILGYTLVLDEN	---EKEKIDN	QYNDMTKLTPLSLEEYKRP	2302			
Db	952	KGADEER	---	HSISELS	SPEEKMLTDNBSGSLN	---	HEEYKEH	990	
QY	2303	QFLMWT	EMAEEFCNRKRDQLK	---	LEAGCKEYBNGSN	-----	2339		
Db	991	TSNSDNYQ	SGGIVNNVMEKEL	QDLENPSSSLD	BGKXHELSSENNLS	SQDMSNTGPL	1050		
QY	2340	DGKTOEAC	EACVYQNF	IKKMTYERQREKKKQ	-----	KDGKTYKDYPISTER	2388		
Db	1051	DNNS	SEETTER	-ISNN	EYVNEBERLTJTKY	EDIVILKSHMNBESD	GELGYLDE	---NS	1104

QY 2622 GCAMPICDITWKYSFADL-----GDIY-----RGTDMLRIGGLPVEIKLYKVE 2667
Db 1259 -----DLQKHGFHTMNNLHGCGVSEBQINSHHGNODRGN----- 1296
QY 2668 YIYKMKNNKNGKRRKYDVOTFRSAMDANKDIWAKMTCAKADAKLF-RKGRMDFER 2726
Db 1297 --SGNVLNMSNNNNNNNT-----PSRYMLYDKKLDLDLXE- 1330
QY 2727 ITLIDKCGHDDPPVDYIQRFRWMTWSEYCKALME-----ELKFKKSCDH- 2777
Db 1331 -----NRNDSTKEEL-----KTLAINKCENETSVKICDHMTHEE 1366
QY 2778 --CITSRCANDIDENKCEOCCTRCOEYKNFVLKMSLFDIOSNRY--KELYEQPIYT 2831
Db 1367 IPIKT--CTREKTRNLCCAVSDYCMST-----FTYDSEBYNCTKRFEDPSYT 1413

RESULT 7

US-09-924-154-17
Sequence 17, Application US/09924154
Patent No. US20020127241A1

GENERAL INFORMATION:

APPLICANT: Narm, David L.
APPLICANT: Sim, Kim L.
TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
FILE REFERENCE: 05213-0465 43170-262105
CURRENT APPLICATION NUMBER: US/09/924,154
CURRENT FILING DATE: 2001-08-07
PRIOR APPLICATION NUMBER: US 60/223,525
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 17
LENGTH: 1501
TYPE: PRF
ORGANISM: Mammalian
US-09-924-154-17

Query Match 3.1%; Score 594.5; DB 10; Length 1501;
Best Local Similarity 21.4%; Pred. No. 1,9e-24;
Matches 331; Conservative 211; Mismatches 529; Indels 477; Gaps 69;

QY 2094 RGGIEGNCNRTYGQYPMGCIYKSKENNGICMPRRKILINNIOYL-NTEENKRN 2152
Db 125 RYIRGCKGR--STHTWIC-----ENKGNNTICIPDRVQICITLQDLKNSGETTRK 178
QY 2153 DIKEAFICAALETQFLMLKTIENPAENELONGTIPDEFKRIMYYTYGDKMEFFGT 2212
Db 179 LTRDVFSAWETDILNKKYGRG-----FDDFCDDYKNSYLDKYVIFGT 226
QY 2213 ISNDKIIIVTSVPTILNENKKKODKKDEELK--IFENKNNKFTMEGTYGLTYHL 2270
Db 227 L--DK-----NNISLVEES--LKRFEKKDSSVNLPTAMWRRTYGRMKMTIQPYA-HL 275
QY 2271 ---IDENEKERIRNYOYNDMTKLTPSLEFVRKRPQFLRPFTEWAEBCN--KRRBOL 2324
Db 276 GCRKPDENE-----PQINRMLLENGKNCRLMKKEKEL 309
QY 2325 KLEAC--KEYCNGSNDGKTOEACAEVYONFIRKMTYERORERFKDKDKKY-- 2380
Db 310 TGECSVNRKKSDCSTG-----CNNECYTRSLI-----NRQRYEVSILGKTYIK 353
QY 2381 -KDYSTERDIEKATCAHEYLNMKLELCKGNKDCSCMKPSSQLPKTYQSOSSDAMP 2439
Db 354 VRYTIFRKRKIYQPNALDFLKLNCSE-----C----- 381
QY 2440 ESLDYVEEFNKCCEPCLSKSGSMITHFKKITEPKIPANCVEKAAYVLSKEANNMDITLK 2499
Db 382 KDIDKRP--FFEEYKGYEEK-----CMQSYIDLKIQKNNDDICSFN 422
QY 2500 EKFTIESTEKESKNSMTNNPNCPKPYVAPDKYIGRRNCENRENRFRVDEYEMCYK 2559
Db 111 1

Db 423 AOTDIYSDKRF-----CLEKKEKP-----WKCDK 448
QY 2560 NSKFYOEKRRVCVPRRREHNCNLNDEI--KIERLDSNYLLKVRRTARRNEGIDIIKN 2616
Db 449 NSEFYHHKGVCSPPRQFCGLNLYLLNDIYVHNSQLLEI--MASKQEGLLMKK 507
QY 2617 ---FSENGCANPICDITMKY--SFADLGIYKRTDMLRIGGLPVEIKLYKVEYI 2669
Db 508 HGTIIDNANC-----KYNDSYVDYKDIYIGNDLMDNNST--KYNNMLNLFERN 557
QY 2670 YGWRNRKNGRNR--YNDVOTFRSAMDANKDIWAKMTCAKADAKLFKGRMDFERT 2728
Db 558 FG---YKVRNKLFTKIELKNNVWLNRKNVWESMR-----GIDEVD 598
QY 2729 ITIDKCGHDDPPVDYIQRFRWMTWSEYCKALMELEKFKKSCCHCTSDCKNDY 2788
Db 599 QRRKTERIDEL--ENMOPFRFMSQWAFPECK-----EKWEYELKINDKCTGN 646
QY 2789 DENKCEQ--CTRCOEYKNFVLKMSLFDIOSNRYKELYEQPIYKISTYHVOYFOKL 2846
Db 647 GKSLODKTCQVNCINMNYWYTRKLAIEIOSVYTK-----DRKL 687
QY 2847 KTFKSECVSEFSSEYLHETSK-CLNYKENENDSSNIFTYAFETPKSYKACSC-----T 2901
Db 688 FSLAKDNVTFE--LKEKAKCSNIDF-----TKIFDQDLKFKRCSGMDYQ 734
QY 2902 LSKNPLNCPDQDKCKELQTFPCSKNDYNNLNNMNAVYLVNSSDDNKGVLLPFR 2961
Db 735 LEVKN-----KEMLSIDSNSBATDISEKGEBELYVNNN----- 769
QY 2962 RRRHLCRPTAYNRKCKEILKILKILTSASFQGLQCKYSEBELCEFAKMSYADYS 3021
Db 770 -----SVSVASGNKEI-----EKSDEKQPEKQKQ----- 795
QY 3022 DIKGTDMNDISLSEKIKIETSNTEATENKRNWENNROIWMALCGYIATSKYTL 3081
Db 796 --TNGTLVRRDKSDRKKGDKDTATD--TKNSP--ENIKVQ-----BHGNGETIK 840
QY 3082 EGMQCPKDEETNOFLRWL--IEMAKQACKERKHVSDSLKTCPSNEDNPFASLQNP 3140
Db 841 EEPRLPSSSETIQSOBELAOKOKEBEPPKKQBEPPKKQBEPPKKQBEPPKKQ 898
QY 3141 CONDRIKTIISLILIKNTMENILIKYKQDKDQSSGNID-----NRPSEENY 3186
Db 899 -----EHEEQKOEEOQIODQSGLDQSSKVGAVASBONNISGQGNV 942
QY 3187 QSYIKSDOCALBLNDINE--IYVGRKNNENNFEKELYKLYGLYFVEDETHKNVLD 3244
Db 943 KSSPEVVPQETTSBNGSSQDTKISTEPENNS--VDRATDSMLDPEKVINENMSD 998
QY 3245 GN-----IKKEEQTVR-----PKALY 3260
Db 999 PNTWTEPDASLKDCKEYDQAKKELQSVSRIBENEDVOSTPREDPTVYGKYGDAEM 1058
QY 3261 FETPHVDSFYAPLPSHRAQYDPKNDILKSSISVIVSAL--GLIALHFMKKKSS 3317
Db 1059 LFSPHAD-----NSESEGLNPPTDI--KTDTGVYKQEBIIGGSASATETSXNLEK 1110
QY 3318 VDLRIINIPQEGYGMPTL-----ESKNRYIYPSGYPKKTYYMGDTPSGDEKTMMD 3372
Db 1111 KDY-----EPHSEISEVYLSGTTGKESELSKSSITTKET-----DPRSNQOEATD 1160
QY 3373 LSSSDITSSSEYEBELDINDIYVGPSPRYTLLEVLEP--SKR-----DIPSDDTPSNDT 3426
Db 1161 VVENSRODNNLSNSVNOQSVNLNREDPLASFEYEVASEPDSRSRIITTEVPSTTVKPPDE 1220
QY 3427 PRTNRFIDDEMELKHPFYQYLPNT--EPNNNYKASADIPN--TEPNTLYSDNPEER 3481
Db 1221 KRSEVEGEKAKEIK--VAPVYPRALGEPMENSVSQSPNVEDVEKETLISENN-- 1273
QY 3482 FIISIHRODLYTGKEISYINIMSTNTNNDIPMARNDSYGIDLINDS 3529
Db 1274 ---GLAH-DTHRG-----NISEKLDLIDHT--LRNEA--GSTIILDOS 1307

RESULT 8
US-09-924-154-14
Sequence 14, Application US/09924154
Patent No. US20020127241A1
GENERAL INFORMATION:
APPLICANT: Narmum, David L.
APPLICANT: Sim, Kim L.
TITLE OF INVENTION: Anti-plasmodium Compositions and Methods of Use
FILE REFERENCE: 05213-0465 43170-262105
CURRENT APPLICATION NUMBER: US/09/924,154
CURRENT FILING DATE: 2001-08-07
PRIOR APPLICATION NUMBER: US 60/223,525
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patent version 3.1
SEQ ID NO 14
LENGTH: 1143
TYPE: PRT
ORGANISM: Mammalian
US-09-924-154-14

Query Match 3.08; Score 591.5; DB 10; Length 1143;
Best Local Similarity 20.88; Pred. No. 1.9e-24;
Matches 293; Conservative 188; Mismatches 453; Indels 473; Gaps 67;

QY 1118 NIIDKYEHEQAQMSVNSGSIASSTAKNHNDRNVEFLSELYOQNGSKNSGTSDE 1177
DB 62 NIYSDVEDVANK-----LINSFVENKSYK-----KRISFL-----NKRKSD- 102
QY 1178 SAVIGMTYEN-----VGAYLHDYGNFDDCQSONEPCD--EKSDGDNERYAFRDKPD 1230
DB 103 --IIPSYSDNFNSISENEDNSGN---TNSNFPANTSEISIGDKNQYFLQK-RT 155
QY 1231 HDAGCGKSGSPRYVQIKTKKABEKDTECKTYNDLIKENDGKQVEDCPKANSNCP 1290
DB 156 HLFACGIR-----KSIKWICRENS----- 175
QY 1291 DMOCINILVEDPRVCMPPRROKLCVHFLANDNEIKKLOSQVNLKEAFKSAAEFPWS 1350
DB 176 -----EKIVCVPRDKRITQICIANFLN-----SRLEPKEKKEIFLISVNTBA--KL 219
QY 1351 YVYSKDEGNELDKELKGIPIAPLRSMPYTFGDRDLFGNDISKHSGSKLKEOI 1410
DB 220 LYNNK-----EGKDPSEIFCNEELNRSFSDFRNSFIDGDMDFG--GNTDRVKGXI 265
QY 1411 DSLF-----KNGDOKSPNGKTRQEMWTHSHETMEAWL-----CALVRIGAKKD 1454
DB 266 NKRTSDYKKEKNVKNL-----NIKKEWKEKKNANLNMHNVHKGNISRECAIIP----- 316
QY 1455 DFTENYGNVYKFSDKSTLEEFAPKRPQFLRLMTEWDDYCYTRQYKLYKDOECKCKSNQ 1514
DB 317 -----AEPOINLMIKEMNENFLMEKKRLPLNFKDCVEMK 353
QY 1515 LK-CDTECNKCEYVYKMKKKEMIP-QDKYYRDER--DKKRF----- 1554
DB 354 YEACFGGGRPLPCSSYSYTFPMKSKTOMEVLTNLVKKKSGVSKNPNFLNDLFKRNKNNDLD 413
QY 1555 -----DROHIGVWYDTGTNATDVLNKRFTASCGDKPCASVY--ORNIOLLEKQAYYAD 1609
DB 414 FFRNEKEEDLDCRCRYATITIKSFLN-----GPAKDVIIASQIVNVL----- 457
QY 1610 KHCCTFIENDRYTNISSKDKCKGLVKEANTGAIKQNKQPNVNNLKELTEDVLEPS 1669
DB 458 RGFEC-VRKSNKSMN-----CTG-----PFTNFPCTCE-----P 489
QY 1670 RLRLICE-----HALDGNTPDEPKDENGIRKRLMEVAATEGINLQYKKEKEKIKISD 1726
DB 490 RRQTLCLGRTYLLHRGH-----EEDYKEHLASTIYEAQLLYKKEKDEML----- 537
QY 1727 AHKSYEVPSCAMKYSFYDLRIILIGIDNLED--EKQTEENLKKIFRNKNTSVGKGS 1785
DB 1727 AHKSYEVPSCAMKYSFYDLRIILIGIDNLED--EKQTEENLKKIFRNKNTSVGKGS 1785

DB 538 -----CSTIQNSYADIADIIKGSILIKDYKKMEENLNKY-NKD---KKRNEE 582
QY 1786 TTGNPGSTARKEFPWNEKCYVNNAMICGKRRGRDNGSGNARSDEDLCKGSPSDDY 1845
DB 583 SL-----KIFREKMDENKENYKMSAVLK-----NKECKYDKFKQIP----- 623
QY 1846 PMGRNREGTAYOFLRWEFMEWEDFC-KHKER--ELEKLVACADNYTCGDNEKRRKCTD 1902
DB 624 -----QFLRWEKEMGDFFCEKREKIXYSFEFYCKKDCDEN-----TCKN 666
QY 1903 ACTQYKFFISMKQOYEQIKRYEKNQ-KIYSEHPAKADAREYLDQK----- 1954
DB 667 KCSSEYKMWIDLKSEYEQVDYKDKKKMYNDIDEVKNRE-ANYVLKESKCKCDVNF 725
QY 1955 --KICENKSGCEYKCMK-----DVSTQRLTDGSGNMPASLD 1990
DB 726 DDKIFNESPNYEDMKCKKDEIKYLNELKYKTKHDIYDIDTFSDTGDGPIISANIN 785
QY 1991 DEP--KEYEGKCNQVPRGPRVRRETPSPVSLISKATA--SKKEAKTAPTPQPKVE 2046
DB 786 EQSGKDTSNSTNGSETSDSP--VSEHESDAIINVEKLSGDESSSETRGDIIDNP--SVT 842
QY 2047 NLITEMRQOTRRRAAOTRRKTS-----TATTESD--VGMVYAILSNKPDNRGI 2097
DB 843 NNVEVHDASNTQGSVNTSDITNGHSESLNRTNADIKIRSGNQSNDQNSHSS 902
QY 2098 EGCNPKTYGYPRKGCIVKSKENENGICMPRRKILCINNIOYLYETENKRRNDIKEA 2157
DB 903 DNGSLRTIGVY-----SEDN-----TQVTVSS----- 925
QY 2158 FIKCALETQFLMLKYLIIENPAENELONGTIPDEFKRIYTYGYKDMFGTIDISND 2217
DB 926 -----QNPBRDTPNALASLPD-----DKINEIEGFRSDS 957
QY 2218 K-----ITYVNSY--TIIENNNKKODKKDELRKIFEKKKFIWEMGITGLTH- 2269
DB 958 ENGRDITTSNTHDVRNIVSERVNSHD-----FTRNMANNNAHQ 1000
QY 2270 -LTD-EN-----EKEIRNDYQYNDMTKLPSEEFYKRPQFLRWFEMAEFCNKRKE 2321
DB 1001 YITQIENNGIIRGDESGNSVNTKD-----NPKSNPSEENDHKKNIOEVSRTKRRRE 1056
QY 2322 QLKLEACKEEYECNG-----SNDGTQPC--AEACVYQNEFK 2358
DB 1057 EIKLS--KONKCNNEYSMECYTSDERNSSPGCSREERKILCCQISDYCLATYFN-- 1111
QY 2359 KMKTEYERQREKFKDKGKRGKDYPS 2385
DB 1112 -YSTY--YNCIRSEIKSPYKCFKS 1134

RESULT 9
US-10-153-273-6
Sequence 6, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Welliams, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:

Query Match.	2.28;	Score 431;	DB 9;	Length 749;
Best Local Similarity	21.5%;	Pred. NO. 7.7e-16;		
Matches 210;	Conservative 128;	Mismatches 262;	Indels 378;	Gaps 44

[illegible]

RESULT 10.
US-09-924-154-16
Coverage 16

```

? Patent NO US20020127241A1
? GENERAL INFORMATION:
? APPLICANT: Natum, David L.
? APPLICANT: Sim, Kim L.
? TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
? FILE REFERENCE: 05213-0465 43170-262105
? CURRENT APPLICATION NUMBER: US/09/924,154
? CURRENT FILING DATE: 2001-08-07
? PRIOR APPLICATION NUMBER: US 60/223,525
? PRIOR FILING DATE: 2000-08-07
? NUMBER OF SEQ ID NOS: 17
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 16
? LENGTH: 972
? TYPE: prt
? ORGANISM: Mammalian
? US-09-924-154-16

```

Query March	2.0%;	Score 382.5;	DB 10;	Length 972;
Best Local Similarity	19.7%;	Ped. No. 5.2e-13;		
Matches 253;	Conservative 157;	Mismatches 450;	Indels 427;	Gaps 49;
QY 1051	LRMWTEAEYCYVQKKETDKLEKCK-ECK-----DKDNGQGTRESGTGTCTKEAC	1103		
Db 16	INMW-----TKINVIEFVQKREICKSCCKKKAMIDKKND-----FTLSEIE	60		
QY 1104	NEYNDITGLTKEQWNIIIS---DKYELHQAQMSVNSGSIASSTAKNHLDRVIEFLSE	1160		
Db 61	LKYNKSSLSLTAKFEYLKQMKDKYELN-----FSSIT	92		
QY 1161	LYQNGKSKSKSGTSDSFAVIGTNTTYENGAVALHDGNGFDDCOSQNEFCDEKSDKQNE	1220		
Db 93	FDDLNNKKYRK-----CICONNKKIENNALYV-----KIRIDICN-----	125		
QY 1221	KYAFRDKPDQHDQAGCGKSGSKPFYVQGLTKTKKAAEKKDTECKTYNDLLKKNQCKKQYEDC	1280		
Db 126	-----NRYKSTYIGBLTKE-----KQND-----	144		
QY 1281	HPKKNSNGTPDMOCGINILVEDPRVCMPPRQKLCVHEFLAND-----NEIKKLQSQVNLR	1335		

Db 145 -----KIMOCINEHIDPEVCGPPRRQOLCGNLDKDEPKVNYNDKFLNEL---192
QY 1336 EAFIKSAAEFTFSWYYSKSGDEGNELDEKELGKIPAPFLBSMFYFDGYRDLFGTD 1395
Db 193 ----ILGIRDSGKFLIEYKRNHMHENMYLDERA-----CKLYNSPFDYKNIILIGKD 240
QY 1396 -----ISKHGEGSKLKEQIDSLFKNGDKSPNGKTRQEWTEHSHEIWEAML 1443
Db 241 MWDPNSIKTENILKGFEEG--IKANIVSMYPSADLSLD-EFRKHMWDONKQOLMEALS 297
QY 1444 CALVKIGAKKDDPLENGVNNVFSODKSTLEEFARQPLRMLEWYDYCYTRQKYLK 1503
Db 298 C-----EFKGNHT-----GVCLEMDNDNDYLMHFRKKNPCJDKLWMD 339
QY 1504 DVOEC-----KSNPOLKCDTECNKCKEDYVYK-MKKKKEMIPQDKYKDERDKR 1553
Db 340 VIKPCIDKKVSKPSENSDVATVONCKSTYDKMILKKREKYNQSSKRYK--RDRSL 397
QY 1554 PDR--OHIGVMTDYGTATNATDYLNKKFTVASCQDKPGSASVVOIRNQLERKAYADAKH 1611
398 FNNVIONL-----KPEYLSMKCT-----E 417
QY 1612 CGCTKFIENDKXYTN-----SSKDKCKGLYKEANTGAIKQNKGNPNYNNLKELEDTYLF 1667
Db 418 CTC-----NLDTOFYVYKGYEDICKSYKp-----444
QY 1668 PSRLRLICFHALDGNATDPEVCEGLRKLMEVAATEGYNLQYKKEKKEKIKTSDA 1727
Db 445 -----YDPEIKDE-----EF 455
QY 1728 HKSYEYVPCGSAMKYSFYDLRIILGIDNLEDEKOTEENLKFIENKNGT-----SVKG 1781
Db 456 NEPSLANNPIS--LISQDTERVSSVDVLSIKENVD--LKPFKPGKQSSHVQVGN 510
QY 1782 GSSTJGNPSTARKFFWNE-----NKECVANNMICGYKRGDRDONS-----GNSARSBDL 1833
Db 511 PRESSESPSANGREDBSTSSSTYNDGVTISSSLGSSSRDVSSTSPVGVDEHEARELL 570
QY 1834 ---KKGVSDDDYPMKRNDEGTAYOFLRWFALWGEDCFKHKELEKLVGACNDYTC 1890
Db 571 POKIIDGVYQSDBSLQSGKSSQEQ-----HNLDGSSLRHSNDOEHSI--ITSQVHE 625
QY 1891 GDNEDRKKCTDCTQYKRTISEKPOYEKOIKYKGNKRTIYSEHPVAADAEAREYLD 1950
Db 626 GTNSLFGSIODET-----ILGESEPLTSP--PEHETSKMDTHAGKRMGEVBNASY 677
QY 1951 KOLKIKICENKSGOEYCKMDVSTQRLTQNSONMPAS-----LDDEPKVEGKQNC 2002
Db 678 DSSSEMSNGRGGLKTKEMGEVYTGITSKNDINLEDSTVHSRONKLENSGDNTOGKEHI 737
QY 2003 QVPRGPRVRERETPPSPVSLISKATASKKEAKTAPTQOPKYEENLTTEKRAOTRTERRA 2062
Db 738 NVLQ-----MDKLENPPTSERDSV--LESESKLNR-----770
QY 2063 QQRKRTSTVTTESDVGYVAKAILSKPDSRGIEGCPNRYGYQPKWCIV-----2115
Db 771 SHVHDNRRIETIAENNIGGLSNSVHGDGRSQRNRHINSRSHGLESIDIVVAGDISN 830
QY 2116 ---GKSEKENGICOMPR-----RKLCINNOYVANYETEKROUNDIEAF--IKCALIET 2166
Db 831 IEGDEEEDANTLKYPRNVLNNKNSRTYNIIEEYIRDVKNVADDIRSYKSNRC-----T 886
QY 2167 QFLMKLIINPAENELONGTIPDEFKRLMYTYVYDGKMFEG-----TDISN 2215
Db 887 NNLSSNVCCK--LKKSLSLNTCTNEDSKRLCISIDYCKMFFPNSSGTHSCMRKREFSN 943
QY 2216 DKTIITVNSVTITLNNKKKODKK 2242
Db 944 HAYKCFAGKGFSSMFLDKKKKKKKK 970
RESULT 11

US-10-153-273-20
; Sequence 20, Application US/10153273
; Patent No. US20020169305A1
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; Chitnis, Chetan
; Miller, Louis H.
; Peterson, David S.
; Su, Xin-zhaun
; Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knodde Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0; Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/153,273
; FILING DATE: 21-May-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,288
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH21.1FMDV1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-10-153-273-20
Query Match 1.98; Score 365; DB 9; Length 411;
Best Local Similarity 31.88; Pred. No. 1,5e-12;
Matches 84; Conservative 17; Mismatches 137; Indels 26; Gaps 5;
QY 858 VIMPPRRHICTSLNEHQTDDHPLNGNIYDDLYVNSFLGDLVLSAKYEANKITRYK 917
Db 51 VFLPPREHICTSLNAXXA-----XAXXXXXXXXXXXXXXXXXXXXXXXXXX 104
QY 918 NNLKGPREVDPKHQTITCAIRYSFADIDITGRQLMERNGDMVLAQHLETVPGNIH 977
Db 105 XXXXXXXXXXXXXXXXXXXXMCBAVRSFADLGIIRGDMDBDKSSXXXXXXXXXXXX 164
QY 978 KSLKGRKNDKYND-----EYLLKLRNWMENARAKYWEAMKCDIKLAKSGHQSQSS 1093
Db 165 XXXXXXXXXXXXXXXXXXXXKYLRLADWEANRHOYWRAMKC-----ATYXX 212
QY 1034 YCGYSDBTPLDDYIPQRLRMTEWAEWYCKYQKREYDKLEKCKECKDKDNGCGCTRESG 1093
Db 213 XCXXXXXX--XKIPQRLRMTEWAEWYCKAQSQBYDKLKKICAXXXXXXXXXXXXXC-XX 268

OY 1094 TGCTKTEACNEYNIDIGLMEQW 1117
Db 269 XXGCKKACADCKYKEIELEKMEQW .292

RESULT 12
US-10-153-273-2
Sequence 2, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Slim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Willems, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESS: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1115 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium vivax
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-153-273-2

Query Match 1.88; Score 343.5; DB 9; Length 1115;
Best Local Similarity 18.8%; Pred. No. 8.9e-11;
Matches 243; Conservative 177; Mismatches 435; Indels 439; Gaps 60;

OY 2304 FLRWTEAEECRNKRKOLLKLEAGCKEYEGNSNDKTOGCAACACTYONFIKKWTE 2363
Db 58 FLHSSQNVNLLIETITLIE---CNEYVKGNGYKLA-K-GHCVEDN-LEK--- 108
OY 2364 YEROREKFKKDKGRKYDYSTERIDIKATCAHEYLMMKLE-----LCGNK----- 2411
Db 109 --LQGTNRSRSEBENIKYK-YGVTELKITYAQMNGKRSRIILKESLYGAHNGGNSYMGK 165
OY 2412 ---DSCNQKSSQLPKTQOSDANDMESLDYVEEFNCKECPPLSKKSGNIHTKK 2468
Db 166 DGGDTGEKGEH--KTDSKTDNGKANLVMIDY-----ETSSNGQ----- 206

OY 2469 IPEPKIIMNCVEKAAAYLSKEAENNDITLKEKFIPIESTKEKESKNSWTNNP--CDPK 2526
Db 207 -----PAGTLNVLEFVYEGHNS-----RNSNGCNPPYDIDHK 241
OY 2527 KEVAP---DKYIGRRNPNCEENREKFKVDYEMKCYKNSKFOEKKVCVPPREIMCLRN 2583
Db 242 KISSAIIHAFLOMTVMNKNVCYKRRRRRDMDC-----NYKKDYICPDRRYOLCKE 294
OY 2584 LDEIKIERLKDSN-----YLKMYRTPARNEGDIITNFENSEGCANP1CDTMR 2633
Db 295 L--TNLVNNTDNTNFHDIYFRKLYLKRKLIDAAVEG--DLLKLN--NYRYKKDCIKR 349
OY 2634 YEFADGDIIVGTMDLRIGGYLPVEIKLYKFEYIYGMKRNKNGKRYNDVOTFRSAW 2693
Db 350 WSLGDFGDIIMGTDMGI--GYSKVVENNRISLF-----GTDE--KAQORRQW 394
OY 2694 WDANRDIWKAMTCAPADAKLFRKGRMDGFERITLIDKCGHKDDPPVDYI----- 2746
Db 395 WNSKAQIWTAMYSV-----KKRLKG-----NFIWICKLVN 426
OY 2747 -----PQFRKMTWSEYICALMELEKFKSCDCHCTSDCKNDYDENK--CE--QCKT 2798
Db 427 AVNIEPQIYRWIREGRDYSELPTVOQKKEKCDG-----KINYDKKCKVPCQN 479
OY 2799 RCOEYKNEFLKWSLFDIOSNKKELYBOPITYKISTYDVQNFQXKLTFRKSEGSSEF 2858
Db 480 ACKSDQMTTRKKNQMDVLSNF-----ISVKAENVOTAGIYTPYDILKOE--LDEF 530
OY 2859 SEYLHETSKCLNFKYKENDGSSNIFTYAFEEPTPKSYKACGCTLPKSNPLDNCPTQNKD 2918
Db 531 NEVAFEN-----EINKRDG-----AYELCYCSV-----EAKK 559
OY 2919 GCKELOTFTFCSKNDYNNLDMNNAVLYLNS-----DDKKGVYLPPIRRRLHCTRPITA 2972
Db 560 NQGEVYT-----NVDMAKSAQATNSNPISQPYDSSKAEKVPDSTH----- 600
OY 2973 YNRRKDEILKKLLITSAFSGQLLGOKYKSEELCFEAMKYSAVDSDIITKGTDMMDT 3032
Db 601 GNVNGGDSSTGKAVT---GDGQ--NGNQTPAESV-----QSDIAESVSAKNV 646
OY 3033 SLSEIKIKIFETSNEATNRTKRTWENNRRQIWMALGCIKATSKVTLDEGWCQLPRDEE 3092
Db 647 DPQKSVSK-----RSD 658
OY 3093 TNOFLRWLEWAKOACKKKHVSLSLTKCPRSNEDNFEASELRLROGCCNDIKYISLN 3152
Db 659 TAS-----VTGLAEKGENLCAANS-----RPSSTVEAN---SPG--DDTVNSASIP 701
OY 3153 ILIKNTMEN-LINIKYKOLKDOSSGNDKRPSEE--NVOYIKSKDSOCALDELINELIVT 3209
Db 702 VV---SGENPLVPTPYNGLRHSKNSDSDGPAESMANPDSNKGKGTGK-----GQDNMDAK 753
OY 3210 GTKNNENNEFEVYKLLPGIYFVEDTHKNHVLIDGNIKKEEQTVRPRALYFTPHVDSF 3269
Db 754 ATKDSSNS-----DGTSSATGDT-----TDAVD-- 777
OY 3270 YQAPLFTFHRVAQYDPKNDILKSSISVYVSAALGLALHFKKKFKSSVDLRLINIPG 3329
Db 778 -----REIN----- 781
OY 3330 EYGAFTLESKRRTYPIYSGPYKGTIYMEGDTSGDED-----KYMDLSSSDIYSS 3382
Db 782 -KGVP--EDRDKTVGSDGGE-----DNSANKDAATVVGEDRIRENSAGSINDR 829
OY 3383 SEYELINDIIVPGSPRYKTLIEVLEPSKRDLPDSDTSNPPRI--NRFDIDENEL 3440
Db 830 SK-NDTEKNAGSTDSQSEDAITALSTESLESTESGDRFTDNTNLEKNGKKEKEDIQ 888
OY 3441 KHDFVSOYLPNTEPNNNYKSAADIPMNTPEPNTLYSDNEEERPFILSHDRDLTYGKEISYN 3500
Db 889 KHDFKSNDTPEEENSD-QTTDAE--GHDRDSIKNDKAEKRRKM-----NKDTFTKNTSHH 942

OY 3501 INNSTNTNDIPNANRDSYRGIDLINDSLVYN 3534
 DB 943 LNSNNLSNG-KLIDIKRYRDYKATREDIILMS 975

RESULT 13

US-09-815-242-12996

Sequence 12996, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-05-23

PRIOR FILING DATE: 2000-05-23

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-10-23

PRIOR FILING DATE: 2000-11-27

PRIOR FILING DATE: 2000-12-22

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

LENGTH: 6281

TYPE: PRF

ORGANISM: Staphylococcus aureus

US-09-815-242-12996

Query Match 1.5%; Score 291; DB 10; Length 6281;
 Best Local Similarity 17.4%; Pred. No. 7.4e-07;
 Matches 713; Conservative 535; Mismatches 1587; Indels 1260; Gaps 173;

11 KGNMAASLEGDAKSPILKESHS-----ARNVLEKYAKNIRHPSKYAKEHYDSEKGL 64
 DB 1974 QVNSAKSALMGDEKLAQAKOTAKSDIGRLDNLNNAORTANAEVDOAPNLAATVAAKNKA 2033
 OY 65 TRKEFGGSPSTPVKNHNYYPYPCNDH-----KEHTNLRKDDVNLKHPGCRGRONFEDD 120
 DB 2034 TSLNTAMG-----NLKHALAEKDNK-----RSVNTDAD 2063
 OY 121 ESEECGNKIRYKRNDAICAPRRHRMCDKN-----LEALNDINTQNIHDLG- 170
 DB 2064 QP-----KQAAVYTAATQAEAITNNAAGSNAENETQVQALNQLN-QAKNLDNG 2110
 OY 171 NVLVTAKEGESIVNNHPRKGTSDACTALARSFADIGDIYRGIDMEKPNVHDEYETGLRE 230
 DB 2111 NKVAOKESAKRALAS--YSNLNNAOSTATSOIDNATVAGVYAAQ-NTANELNTAMGQ 2167
 OY 231 VFKEHIDGMEDEYKNDYN-----PDGSGNYY-----KLEAMWNNVNRK 269
 DB 2168 LONGIND--ONTYKQOQVNTDADQKKADATTAATNNAOGILDKRAHGQNMTKAQVEALNQ 2225
 OY 270 VW---EATITDASYKSGYFMOSESNTPLFSNPKCGHKQ-----GKVPNT----- 310
 DB 2226 VTTAKNMLNDANVRQAK--SDAKANLGTLLHNNNAOKODLSIQEGITVNGVNGVYTKA 2284

OY 311 --LDVYPOILRMFDEWGEERPCRRNITKLKVDSCNDKDEBLYCSINGHDCITTTTKKGI 368
 DB 2285 QDLDGAMQRLQ-----SALANKDOTKASBNYIDADPT--KK-- 2318
 OY 369 LHLDNKCTDSCSTKCVKEVYEWLGNQOAPFKKOKKEKEKESIOSYLSNKRFPVNNINSEYKQ 428
 DB 2319 TAFDNRITQA-----ESTL-NKDHGANKKOAVEOAIOSTSTENALNGDNLNLO----- 2366
 OY 429 FFEKLETOYATNDPFLNL-----LNEGKYCKGGLPGEKDIPTFSNADKGIYFYS 479
 DB 2367 -RAKTEAIGADIDLNLTHLNTPOKFTALKQOVNAQGRVSGVTDLKNKASATSLNA----- 2416
 OY 480 ETCQVCPDGCWKCDGIKYTHKSDNDRERVNNEDEYKPPGKVPNTITVLYSGNEGDIYOK 539
 DB 2417 -----MDQLKQALADHD-----TIYASG----- 2434
 OY 540 LEMFGSSTNYKDKNNQKMECYKNDINRCKLEQNTFLINDPKIISHNFEFLAVTYL 599
 DB 2435 --NYTNAS--PDKQAVYDAYNAKN-----IYNGSPVIT--NAADVTA-- 2474
 OY 600 LBDTIKMNDKLKTCTINNTTTHCIDECNRNCLCFDRVWVKOKKEEWSIKKLTFRKKNIQOS 659
 DB 2475 -----IQRVNNAETGLNGDIN-----LATAKQAKKALQRMTHLSDAQOS 2515
 OY 660 YSNINNLFEGETFFKYMDKLDKDEAKWKLMEKIKKKKEFSLENNRDYLENAILELID 719
 DB 2516 ITGQIDSATO-----YGVQSVKSNATMLDAMNOLRMSI--AND 2554
 OY 720 HKEET-----ATICKNNTEA--CETSHNATNPQVPRGSTOPTKNKELAQFKRS 771
 DB 2555 DVASOPVYDADRDKONANITAVTAENITINATISQPTLDSAVTAQANOVST-----NKT 2609
 OY 772 AVEEARNRGLHLKGAHEGIYKRGGRKDKDNLIRIMIKHSNR-----NLGFSNGPC 825
 DB 2610 ALNGAQNLANKKQETTANIN-----QSLHNNNAQKODLNQVYNAP- 2650
 OY 826 DKGSTGDIOT-----REVVGTEWVDEPHMKKHEDVIYMPRRRHITSNLEHL 875
 DB 2651 -NISTVNOVYTKAEOLDQAMERLINGIQ--DKDOVKOSVNTDADPEKOTAVNNAVTA 2706
 OY 876 QTDHPLNGNIVDDLVNNSIFGVLVSAYEANKIIRMYKEKNLKGPRKVEYTPKQOT- 933
 DB 2707 ENIIINOANGTNANOSOEALSTVTTT-----KALNGDKRVYDAKNNANO 2752
 OY 934 TICRAIRSEAD-----IGDIIRGRDLMERNGDVKLQGHLETFVGNIRHSLKQ-----GN 985
 DB 2753 TLTSLDNLNNAQGAATGNTINQAHVTAEYV-QAIOGAQELNTAMGNLKSNDKDTLIS 2811
 OY 986 DKYNDAPKYTLKLEKMWME-----ANRAKVEAMKCDIK 1019
 DB 2812 QNFADADPE-----KKNAYNEAVHNAENILNKSTGTVPKQDVEAAMNOVATKAALNGTO 2867
 OY 1020 YLKDKSGHOSTOSSGYGSHPLD--DYIPQKLRMTTMAEWKCYOKKE-----YDKLK 1073
 DB 2868 NLEKAKOHANT--AIDGLSLITNAQKEALKQLOOSTYVAEADGNEBKANNVDAAMDKLR 2925
 OY 1074 EKCECKDKDNGOGCTKESGCTKCTEACNEBYNDIIGLKEQMOINIISDYKELHQA-- 1131
 DB 2926 QSLIAD-----NATTKQN-----QNTDASQNKKAYNNAVYTAQOIIQOTIS 2967
 OY 1132 ---QMSYNSGIEASSTAKHIDRNVIEFLSELYQONGKSKSGTSDESAVIGNTTYE 1188
 DB 2968 PTLDPVYINQAOQAVSTTKKALNGN--ENLEAAKQO--ASQISGLSDNLNNAOKOTVTD 3022
 OY 1189 NV-GAYLHDTGNFDQOSQONEPDEKSDGKDNKRYAPRDKPOHDGACGCKSSKPTPVQ 1247
 DB 3023 QINGAHIVDEANOIKONAQN-----LNTAMGNLKQALADK-----DAIKAT 3063
 OY 1248 IKTRKRAEKEDECTKYV-----NDIILEKNDGK-----KQVEDCHPKKNSNGYPDW 1292
 DB 3064 VNTDADQAKQOQAVNTAVTAENITISKANGNMTQAEVEGAIQOVNAAKQALNGN----- 3118
 OY 1293 QCGNINLVEDPRVCAKMPRRQKLCVHFPLANDNETIKLOSOVNLKEAFTKSAAEFTFSWY 1352

Db 3119 --ANVOHAKDEAFAL-----INSSNDLNOAQKDA-LKQ---QVONATTVAGVNN 3161
QY 1353 YKSDGEGNELDEKELKGIIPAFILSMFYTFGQYRDLFGTDLISGHGSGSLKRLQIDS 1412
Db 3162 VKQTAQELNNAQMOLKQ-----GJADEQYTKADG 3190
QY 1413 LFKNGDGKSPNGKTRQEMWTEHSHEIWEA-----MICALVKIGAKKDF 1456
Db 3191 NFVNADPDKONA-----YNQAVAKABALISATPDVVPVPSSELTALNKTYQAKNDL 3241
QY 1457 TENYGV-----NNYKFS-DKSTLEEFARQFRLMTE-----WYDDCYTRQKYK 1503
Db 3242 NGNTNATKAKONVOHAIIDOLPLNO--AQDEYSKQITQATVYPNVA100AATLMDAMT 3300
QY 1504 DVOEKCSNDOLK-----CDTECNKCEDYKYMKKKKKEMIPQDQYKDEDEKRPFR 1556
Db 3301 QLKQJLANKQIKSGSEYHADYD--KQATYONATYKAEELKQ-----TNTPTMDP 3350
QY 1557 QHIGVAVTYTGYN-----ATDYLN-----RKFTASCGDKPGSAS 1591
Db 3351 NTLQALTKVNDTNQALNGNOKLADAKODAKTTLGLTDLHNDLNOAQKALFTQVQAPDIAT 3410
QY 1592 V--VORNIOLLE-----KOAYYDADKHGCTKEIEND-----DKYNNISKDKCKLVE 1639
Db 3411 VNNVKNQONLNNAMTNLNNALNADQKTETLSINFETDADAKKDAYTNVVS--HAEGLISK 3468
QY 1640 ANTGAIRKQONKGPNNYNNIKELTEDVLFPSSRLICFHALDNGYTDPEVKEDEGLKRLM 1699
Db 3469 AN-----GSMASQT-----EVEQAMQRYNEAKQALNGNDNQYRAKD-----AA 3506
QY 1700 EVAATEGYNLQYKKE--KKEKEKITSAPHKYSTEVPSCSMKSYFDLRIIGIDNL 1757
Db 3507 KOYTNNANLNOAQKALQOVDAAQT-----VAVNTIKQTAQDLNOMATOLKQ 3557
QY 1758 EDEKQTEENLKKI-----FNKNGSYGKSGSDSTGNGPSARKFEMENKCYWMMN 1810
Db 3558 IADKQOTKANGFNVANADTDKQANNNVAHAQIISGTPNA-----NVDQOYAQALQ 3610
QY 1811 TCGYKRGDQSGNSARSDEDLKKGVSPPDDYPMKNGRDEGTAYOLFMPAE--MGE 1868
Db 3611 QVNOAKGJLNGHNLQVAKDNANTAIIDOLPN-----LNGQKALQKQVSHAEVLGV 3663
QY 1869 DECKHEKELELVGACN-----DYTCGDNEDKCKCTDQCTQYKKTISEK 1915
Db 3664 NAIKQNALNNAAGTLKQOIOANSQVPSVDTQAD--ODKQOAVNMANAQO--IANGI 3721
QY 1916 P-----QYEQKIRKYGKNDKIYSEHPVADADAREYLD-----KOLKICENKSG 1962
Db 3722 PPTVLTPTVTOAVTMMQAKALNGDEKLAQKQALANLDTLMDLQPODALRLNOIN 3781
QY 1963 DCEYKCMKDVSTQRLTDSQNMPSLDEPREVEGKCNCOVPRGPPRYRRETPSPRYSL 2022
Db 3782 QAO-----ALATVEQTKQNAQNVNTAMSNLKGINKDTYKASENYHADADKQAYTNA 3836
QY 2023 ISKATASKKEAKTAPTKOPKKEVEMLTJEMRQOTRRAAQOTRKRRTSTATTESDVCTM 2082
Db 3837 VSQAGIINO--TTPPTLANPDEITALTVO--TDANKLNGEALVATEKONAKAQAVSQM 3891
QY 2083 VKAILSNPDSRSGIEGNCPTTYGQYPMWGCIVGSKSENGICMPPRKIKICINNI--- 2139
Db 3892 THLNDQAKQALKGQID-----QSPET--ATVNGQVKTATSLDQAMQLSQALNDRKQT 3942
QY 2140 -----QYLVETENKRDNDIKAFIKCALETOFLMKYIINPAENLQNGTTPDEKFR 2195
Db 3943 LADGNYLVADPD--KONAYKQAVAKAE-----LNLKQSGTNEVQ----- 3980
QY 2196 IMYVYGDYKMFQGTDISNCKIITVNSVTIINENKCKOKKDEDEKRLKTFWEK 2255
Db 3981 -----AOVESTNEVNAKQALNGDNL-----ANA 4006
QY 2256 KFIWEGMIGLYTLHJTD--ENEKIKRDNYQYNDMTKLPSLEEFYKRPQFLMFTEMA 2312
Db 4007 KOQAKQOLANLT--HLNDQAKQSESQITQAPLYTDTVTINQAKQTLDMAMELR----- 4059
QY 2313 EEFCKRKQOLKLEAGCKEYCNQSGNDKTOECABACTYONETIKMKTEYER----- 2366
Db 4060 -----NSVADNOTTLASE--DYHADATQORQNDYQAVTAAN 4093
QY 2367 -----OREKFKDKQKTKKDYPTSTPRDEKATQCAHEYANMLK 2405
Db 4094 NIINQSTSTPMNPDVQNCATTOVNNTKVALDGD-----ENLAAKQOANNRILD 4141
QY 2406 ELGKNKDCSQMCKPSSQLEKTYQOOSQSSDANDMPESLDYVPEEFNK-----CECPELSK 2460
Db 4142 QL--DHILNNAKQOOLQSOI--TQSSDIAVNGHKQATASLNTAMGNLNLATADQAVQOR 4197
QY 2461 GSMITHKTEPKIIMN--CYEKAATYLSKEANNDITLKEFT----- 2503
Db 4198 GNFILAND--TDKQATYNTAVNNAAMINKQOGQANQREVEQATLKQYTTQALNGDHL 4255
QY 2504 PIESTKEKESKSNMTNNPCDPKRPYAP----- 2532
Db 4256 QVAKTNATQALDALSLN--DPQKALMDQYTAATLYTAVHIOEQNANTLNOAMHGLRQS 4313
QY 2533 -----KYGRNPNCEENRENRK-----VDYEMKCYKSKFYOEKKRYCVPRR 2576
Db 4314 IQDNATKANSKYINEDQPEQOONYDOVAOANNIINQETATLDNNAIQAAATVTNTKAA 4373
QY 2577 EHMCLRNDEKIERLKD-----SNLKMVRTANEGIDIKN 2616
Db 4374 LH-----GDVKLQNDKQKHAQVVSQALHNNNAOKHMEDTLIDSETTETAVAKODL----- 4422
QY 2617 FNSENGCMANPDIQPMKTSFADLGIYRGDMLRIGLYLPEVEIKLYVFEYIYGKMRNK 2676
Db 4423 --TEQAQLDQLMALQOSIAD--KDATPSS-----AYVAE--P 4456
QY 2677 NKGRKRYNDVOTFNSAWMDANRDKIWKAMCKADAKLEFRGMDFERITLLQDK-- 2733
Db 4457 NKQSYDEAVONASISLIGLNNPTINKGVNSAQO--AVISSKNAIDYVER--LAQDKQTA 4513
QY 2734 --CGHKDD--PPYDDYLPORFRMNTMSEYKCK-----ALMELEKFKKSCDCKTSR 2783
Db 4514 GNSLNHLIDLTPAQQALQENQINNATTRGEVAKQTLERQALQNMQEMALRNSIQDOQTEA 4573
QY 2784 CKNDYENKCBQCTRCQOEYKNFVLKMSLEFDIOSNKYKELYEOPITYKISTYDHYONFV 2843
Db 4574 GSKFINEBKPO-----KDAYQAVQNAKDLING--TNPPTLDKQ--V 4612
QY 2844 OKLTFFKSECSVESFEYLHETSKLYNK-----FENDQSSNIRKYATE----- 2888
Db 4613 EQLTQ-----AVNOAKDNLHGDQKLADKQHAVYTDLQNLGNLNPORQALESQIINNAATR 4667
QY 2889 -ETPKSYKEACSTLPSKNPLDNCPTD--ONKDKCELEQTFTECSKNDYNNLNDMMAYL 2945
Db 4668 GEVAKQLEAKALD--QAMQALRNSIQDOOOTESKTI-----NEDKPOKDAYOA-A 4717
QY 2946 VLNSD-----DNKGVLIPRRRLCTRPITAYNRKDKIELK----- 2984
Db 4718 VONAKDLINQGNPTLDRKSQVEQLTQAVTAKDNLHGDQKLARDQQAQVTVYNAALPLNH 4777
QY 2985 --KLLLSAFSQGL--LQOKYSEBEL--CEPAMK-----YSTA----- 3018
Db 4778 AQOQALTDALNAAFTREVAOHQVOTATELDHMETLKNKYDOVNTDKAOPNYTEASTDK 4837
QY 3019 -----DYSDLIKGDMMDTSLSEKIKKIFETSNQATENR-----KTMWENNRQ 3062
Db 4838 EAVDQALQANESITDPTNGSNANKDANDVQLYTLQIEKENLNGNERVAEAKTQAKQOTIDQ 4897
QY 3063 ITHAMLCGYIATSKVTLDGWCQLPRDEETNOF--LRWLIEAKQACEKKHVSLSLKT 3120
Db 4898 LTH--LNAQDIATAKQONIDATKLOPIAELVDAQATQUNQSMQDLQAVNHAVNEQFVY 4955
QY 3121 KCPRSNEDN-----FPASLLRQPGQNDIRKYISANILI-----K 3156
Db 4956 TOADSDKQNAKYKALIDAEVNLKQNNANKQOVDQALO--NIIINAQALNGDERVALAKTNGK 5014

Db 940 KIRNVOLESLNHEK-GK-----LQREVDRTIRAHVAEKNIOHLSQHSFDEK 991
 QY 2412 DCS-----CMQRP-----SSQLEPTTQO-----SOSDANDMESIDYVEEFNK-C-ECPBLS 2458
 Db 992 ELEHLOJCOGRSHLQKQFESHQLOLQNKAKENNDKIQRLN---ELEKSEMECML 1048
 QY 2459 KKSMTITKITEPKITPMCEKAAVYLSKEAENNDITLKEKPIESTREKESKSWT 2518
 Db 1049 KQVEELTRONNETKLMQRIQ-----AESEN---IVLEKQ-----T 1082
 QY 2519 NNNDOPKRYAPDXVYGRNPPCENRENFVYEMKCKNSFVOEKKRVCPRRH 2578
 Db 1083 IQORCEALKIQAD-----GFKDQLRSTNEH-----LHKQTEBDF-----QRKI 1122
 QY 2379 MCLRN-----LDEKIERLSDNYLLKAVRTANNEGIDILKNPENGSCAMPICD 2630
 Db 1123 KCLEEDLAKSONLVSEK-QKCOQOIIIONTKKEVNLNAEL--NASKKEKRGQKVQ 1179
 2631 TMYKSPADLD-IVRGTDMLRIGGYLPVVEIKLYKVEYIYGKWRNKNKGRK- 2682
 1180 LQQAQOVELNNRRLKQYODELHLKTEIEQMTIRKKVYLFQESGKFKQSAEPRKMEKME 1239
 QY 2683-----YNDVQTFRSAMMDANKRDKMTKAPEDATL-----FRKGRMG 2723
 Db 1240 SKVTENDIGIRLDFVSLQOEN-----SRADENALCETINIKELEROLQOYREOMQOG 1293
 QY 2724 FERITLIQDCGHRDPPV-----DYIPQFRMTMSEYCYKALMELEKFKSCDHC 2778
 Db 1294 QHMEANHYQCKQKDELDIAQKREVENLKQKMOQIKENHEQVLVLOCETIK-KSTAKDC 1352
 QY 2779 KTSDRCKNDYENKCEQCKTRCOEY-----KNFVKKWSLFDIQSNRYKE 2823
 Db 1353-----TFKPDF-----EMTVKECOHSGELSSRMTGHLPTPRSPILRMTQ---EPOPLEE 1399
 QY 2824 LYEQPITKISTYDHOVNFQKLTFFKSECSVESSEYELHETSKLNYKFNENDSS--- 2880
 Db 1400 KMOHRAVEQIP-KEVO-FOPPGAPLEKESQOCYSEFSOTSTELQITFDETNPITRLS 1456
 QY 2881-----NIRTYAHEET--PKSYKE-ACSTYL-----PSKNPLD 2909
 Db 1457 ELEKIRQALNNSRPVRYQDNCEMELVYLPLEIAKKNQYDMHTEVTTLKQKNPVP 1516
 QY 2910 NCPPTDKKDCKE---LQTFPCKNDYDNNLNNMAYLVINSDDN----- 2953
 Db 1517 SA-BEMMEGCRASGKKGDFLKGLEPETFQNFDDHACSVDDEKFGGLRHVTAR 1575
 2954---KGVLIIPR-----RRHICTRITAYNRYKGDKE----- 2981
 Db 1576 QLVAKLLDMTTIRQLRLGLKTVAEVOKTLNKFITKATSTAGLYLESTKEKISFASAER 1635
 QY 2982 ILKRLLTSAFSQO-----LLGOKYKSEEL-----CFEAMK-----YS 3016
 Db 1636 IIIDKMWALAFLEQAATGFIIDPISQOTYSVEDAVLKGVYDPEFRIRLLEAKAAYGYS 1695
 QY 3017 YADYSIIKGTDMADTSLSEKIKIIFETSNEAETENRKTWMENNRQIMHAMLCGYKTATS 3076
 Db 1696 Y-----SKTILSVF-----QAMENRML---DROKGHILEA---QIASG 1728
 QY 3077 KYTLEDEGMCOIPKDEETNOFLRMLIEMAKQACEKKHVSLSLTKCPRSNEDNFEA---S 3133
 Db 1729 GVIDPVRGIRPPPIALQOGL-----LNNAILQFLHEPSSNTRVPF--NPNNKQALYYS 1780
 QY 3134 ELLEPOPCONDIRKYISLNLIKNTMENLNIKYQLDKDQSSGNDNRP-SEENQSYTKS 3192
 Db 1781 ELLEM--CYFDVESQCLPFGERNISNLNK-----KTHRISVDYTKGSELTYE---- 1830
 QY 3193 KDSOCALENDINEIYGTAKNENNEFEVKLKLIPGLIYVEDETHKNHVLG----- 3245
 Db 1831-----AFORMLIERISYIELSGOQOYQWKEAM-----FFESYGHSSHMLTDTKTGLHF 1877
 QY 3246 NIKEE-EOJVRPKALYFTFPHVDSFYQAPLESTHRA-----QYDPKNDLKSSISVIV 3299

Db 1878 NINEAIGOTTIDKALV-----KKYQGLITLTELADSLLSRLVPKK-LHSEVAGYWL 1929
 QY 3300 SALG--LIALHFMKKREKSSVDLRLINIPQEGYMPLESKNRYIPYRSGPYKGYIY 3357
 Db 1930 TASGERISVILKASBRNLVDRTILRLCLDAQYSTGI-----IDPLTKKRYV 1976
 QY 3358 MEGTSGDEKRYMDLSSDITSESEYEELDINDIYVPSPKYKTLIEV 3408
 Db 1977 AVALHRLVDEGF-----AQOLRQCELYITIGHPITNKKMSYVEAV 2018

Search completed: April 28, 2003, 10:49:06
 Job time : 156.127 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:46 ; Search time 8.41989 Seconds
(without alignments)
4367.892 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554
Perfect score: 1543
Sequence: 1 DCHPKNSNGYPMOCGNIN.....KKEWIPDKRYVDERDKRRF 276

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

al number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154.3	100.0	22	AAB62142	P. falciparum FCR3
2	634.5	41.1	294	AAB62147	P. falciparum varc
3	624	40.4	2703	AA70236	P. falciparum varc
4	624	40.4	2710	AAW22482	Plasmodium Proj3
5	624	40.4	2710	AA77904	P. falciparum Proj
6	624	40.4	3060	AAW22475	Plasmodium var-7
7	624	40.4	3060	AA77905	Plasmodium var-7
8	609	39.5	407	AAB62148	P. falciparum varc
9	524	34.0	351	AAB62149	P. falciparum varc
10	516.5	33.5	308	AAB62151	P. falciparum varc

11	493.5	32.0	793	AA70234	P. falciparum E31a
12	493.5	32.0	921	AAW22480	Plasmodium E31a
13	493.5	32.0	921	AA77902	P. falciparum ebl-1
14	486	31.5	311	AAW22480	P. falciparum varc
15	390.5	25.3	2228	AAW93944	P. falciparum PFEW
16	382.5	24.8	2182	AAW22476	Plasmodium var-1
17	382.5	24.8	2182	AAW77906	Plasmodium var-1
18	359.5	23.3	2913	AAW00384	Plasmodium falcip
19	339.5	22.0	2197	AAW18352	Plasmodium falcip
20	296	19.2	1726	AAW00385	Truncated Plasmodi
21	289.5	18.8	1143	AAW76759	Plasmodium falcip
22	282	18.3	972	AAW76761	Plasmodium falcip
23	274	17.8	350	AAW07656	P. falciparum EBA-
24	274	17.8	1435	AAW70232	P. falciparum SABB
25	274	17.8	1435	AAW22477	Silicic acid bindin
26	274	17.8	1435	AAW77900	P. falciparum SABB
27	274	17.8	1604	AAW70105	TNF-R-EBA 175 fus
28	274	17.8	1786	AAW41043	CD4-EBA175 fusion
29	273.5	17.7	311	AAW77915	Plasmodium DBL gen
30	261	16.9	1421	AAW76764	Plasmodium DBL gen
31	260	16.9	616	AAW50533	Undifferentiated ambo
32	260	16.9	749	AAW70233	P. falciparum EBL-
33	260	16.9	749	AAW22479	Plasmodium ebl-1
34	260	16.9	749	AAW77901	P. falciparum ebl-
35	260	16.9	1086	AAW76760	Plasmodium falcip
36	248.5	16.1	1700	AAW18144	Plasmodium falcip
37	231	15.0	302	AAW07655	P. falciparum EBA-
38	223	14.5	1028	AAW07654	P. vivax PVDBP pro
39	223	14.5	329	AAW41044	Plasmodium vivax P
40	223	14.5	1061	AAW70231	P. vivax DABP
41	223	14.5	1115	AAW13457	Duffy receptor
42	223	14.5	1115	AAW22478	Duffy antigen bind
43	223	14.5	1115	AAW77989	P. vivax DABP bind
44	221.5	14.4	1245	AAW70106	TNF-R-PL. vivax Du
45	218.5	14.2	324	AAW77911	Plasmodium DBL gen

ALIGNMENTS

RESULT 1	ID	AA62142	standard; Protein: 3542 AA.
XX	AA62142		
XX	AC	AA62142	
XX	DE	29-MAY-2001 (first entry)	
XX	DE	P. falciparum FCR3, varCSA protein.	
XX	KW	FCR3, varCSA protein; chondroitin sulfate A; CSA; var gene; PFEWPI;	
XX	KW	erythrocyte membrane protein 1; parasitized red blood cell; PBC;	
XX	KW	malaria; protozoacide.	
OS	XX	Plasmodium falciparum.	
XX	PN	WC200116326-A2.	
XX	PD	08-MAR-2001.	
XX	PF	01-SEP-2000; 2000WO-US24195.	
XX	PR	01-SEP-1999; 99DS-0152023.	
XX	PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
XX	PI	Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;	
XX	PI	Gysin J, Pouvelle B, Fujii N, Smith J;	
XX	DR	WPI; 2001-235109/24.	
XX	DR	N-PSDB; AAF57301.	
XX	XX	Novel FCR3, varCSA protein, useful for modulating parasitized red blood	

PT cell binding, sequestration and onset of maternal malaria -
 XX
 PS Claim 12; Page 63-71; 78pp; English.
 CC
 CC The invention relates to a P. falciparum FCRI3 varCSA protein, that is
 CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
 CC corresponding P. falciparum erythrocyte membrane protein 1 (PFEMP1)
 CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
 CC protein and the encoding gene are useful for treating and preventing
 CC maternal malaria in a patient identified at a risk for contracting
 CC maternal malaria or in a patient afflicted with maternal malaria. The
 CC present sequence represents the P. falciparum FCRI3 varCSA protein.
 CC
 SQ Sequence 3542 AA;
 XX
 Query Match 100.0%; Score 1543; DB 22; Length 3542;
 Best Local Similarity 100.0%; Pred. No. 8.5e-138;
 Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 DCHPKKNSNGYPMOCGNINLVEDPRVCMPPRRKQKLCVHFLANDNEITKLQSOVNLKEAF 60
 1279 DCHPKKNSNGYPMOCGNINLVEDPRVCMPPRRKQKLCVHFLANDNEITKLQSOVNLKEAF 1338
 61 IKSAAAETFFSWYKSKDEGNELDEKELGKIPPAFLSMFYTGDRDLEFGDISK 120
 1339 IKSAAAETFFSWYKSKDEGNELDEKELGKIPPAFLSMFYTGDRDLEFGDISK 1398
 121 GHGEGSKLKEQIDSLFRNGDQKSPNGKTRQEWMTESHEIWEAMCALVKIGAKKDDFTE 180
 1399 GHGEGSKLKEQIDSLFRNGDQKSPNGKTRQEWMTESHEIWEAMCALVKIGAKKDDFTE 1458
 181 NYGYNVNFSDKSTLEEFKRRQFLRLLEWYDDCYTRQKYLKVOEKCKSNLDLQCD 240
 1459 NYGYNVNFSDKSTLEEFKRRQFLRLLEWYDDCYTRQKYLKVOEKCKSNLDLQCD 1518
 241 TBCNKKCEDYVYKMKKKEMIPQDKYKDERDKRRF 276
 1519 TBCNKKCEDYVYKMKKKEMIPQDKYKDERDKRRF 1554
 DB
 RESULT 2
 AAB62147
 ID AAB62147 standard; peptide; 294 AA.
 XX
 AC AAB62147;
 XX
 DT 29-MAY-2001 (first entry)
 XX
 P. falciparum varCSA polypeptide R29DBL2-gamma.
 FCRI3 varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
 erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
 malaria; protozoicide; R29DBL2-gamma.
 KM
 KW Plasmodium falciparum.
 XX
 OS WO200116326-A2.
 XX
 PN 08-MAR-2001.
 XX
 PD 01-SEP-2000; 2000WO-US24195.
 XX
 PF 01-SEP-1999; 990S-0152023.
 XX
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
 PI Gysin J, Pouvelle B, Fujii N, Smith J;
 PI WPI; 2001-235109/24.
 DR Novel FCRI3 varCSA protein, useful for modulating parasitized red blood
 PT cell binding, sequestration and onset of maternal malaria -

XX
 PS Disclosure; Page 72; 78pp; English.
 XX
 CC The invention relates to a P. falciparum FCRI3 varCSA protein, that is
 CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
 CC corresponding P. falciparum erythrocyte membrane protein 1 (PFEMP1)
 CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
 CC protein and the encoding gene are useful for treating and preventing
 CC maternal malaria in a patient identified at a risk for contracting
 CC maternal malaria or in a patient afflicted with maternal malaria. The
 CC present sequence represents a P. falciparum varCSA polypeptide
 CC R29DBL2-gamma.
 CC
 SQ Sequence 294 AA;
 XX
 Query Match 41.1%; Score 634.5; DB 22; Length 294;
 Best Local Similarity 45.5%; Pred. No. 2.5e-52;
 Matches 131; Conservative 37; Mismatches 91; Indels 29; Gaps 10;
 1 DCHPKKNSNGYPMOCGNINLVEDPRVCMPPRRKQKLCVHFLANDNEITKLQSOVNLKEA 59
 12 NCNANKRKN--EWQCDKNTFVGNEGVCMPPRRKQKLCVHFLANDNEITKLQSOVNLKEA 66
 60 FKSAAAETFFSWYKSKDEGNELDEKELGKIPPAFLSMFYTGDRDLEFGDISK 119
 67 FKCAKAKETNLMQKYNKNDKNAEEL--LKKGIPEDFMIMFYTGDRDLEFGDISK 123
 120 KGHGSKLKEQIDSLFRNGDQKSPNGKTRQEWMTESHEIWEAMCALVKIGAKKDDFTE 174
 124 K--DVDRYKKNKINVFNNSSKRGFKIDPENWNNENGPQIWNQMLCALIHADKDSIKN 180
 175 KDDFTENYGNVNFSDKSTLEEFKRRQFLRLLEWYDDCYTRQKYLKVOE 229
 181 KD---NKYKQVITLAKRQDSNGKTLSEFAKKRFLRWFVWYDDCYTRQKYLKVOE 236
 230 KCKSND--QKCDTECNKCEDYVYKMKKK--EWIPQDKYKDERDKRRF 274
 237 TCKSIDGQLKCDRGCNNKCEYKRYKRRKKKEENLQDKYKDERDKRRF 284
 DB
 RESULT 3
 AAR70236
 ID AAR70236 standard; protein; 2703 AA.
 XX
 AC AAR70236;
 XX
 DT 22-SEP-1995 (first entry)
 XX
 DE P. falciparum Pro3.
 XX
 KM Erythrocyte binding ligand; Pro3; binding domain; malaria; therapy;
 KW vaccine.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO9507353-A.
 XX
 PD 16-MAR-1995.
 XX
 PF 07-SEP-1994; 94WO-US10230.
 XX
 PR 10-SEP-1993; 93US-0119677.
 XX
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Chitnis C, Miller LH, Peterson DS, Slim KL, Su X;
 PI Wellens TE;
 PI WPI; 1995-123427/16.
 DR N-PSDB; AAQ83529.
 DR New erythrocyte binding domain polypeptide(s) - isolated from
 PT Plasmodium binding proteins, used in diagnosis, treatment and

Pt	Prevention of malaria
Xx	
Ps	Disclosure; Page 61-65; 81pp; English.
Cc	Erythrocyte binding ligand (EBL) family genes were cloned from
Cc	P. falciparum chromosome 7 subsegment libraries constructed during
Cc	genetic studies of the chloroquine resistance locus. The 4 genes,
Cc	EBL-rel (AA083526), E3la (AA083527), EBL-e2 (AA083528) and Proj3
Cc	(AA083529), encode the proteins given in AAR70233-36, respectively. The
Cc	binding domains of such proteins can be expressed e.g. in E. coli,
Cc	yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
Cc	cells, and provide protection against P. falciparum.
Xx	
Sq	Sequence 2703 AA:
Query Match	40.4%; Score 624; DB 16; Length 2703;
Best Local Similarity	43.8%; Pred. No. 4.6e-50;
Matches 128;	Conservative 49; Mismatches 83; Indels 32; Gaps 11;
Oy	1 DCHPEKNSNGVPMQOC-GNINLVEDPRYCMPPRRCKLCYHFLANDNEIKKLSQVNLKEA 59 :: :: 1360 ECNPES---PDMDCNNNDISHD-GACMPRROKLCLYLAIHSQTEINTDMLKA 1415
Dd	60 FIKSAAEFTFSWYYKKSK-DDEGNELKELKEGKIIPAFILRSMEFYTGDRDELGTDI 118 1416 FIKTAAETFLSWQYTKRKNSSEAILDR----GLPSOLFSLSMATTFEGRIDICNTDI 1471
Oy	119 SKHGEGSKLEEQIDLSLFENGSDOKSPNQKTROEMWTSHSETWEAMCALVRIGAKKD- 177 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : 1472 SKKODVAKAKDKICKFESKDGSRKSPSGLSRGEMWKTNGETIMKGMLCATKYVTDTDNK 1531
Oy	178 --FTENGYCINNVKSFDSK-STILEEFARKPOFLRLTENTDYCYTROKTLKVQEKCS- 233 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : 1532 RKIRNDYVDKRNOSQNCPNPLSEEFNAKPQLRMVIEMGECECAEROKKENIHKACNEI 1591
Oy	234 NDOLKCDP--ECNKCEDYVKMK-KRKEWI-----PODKYYK 268 : : : : : : : : : : : : : : : : : Db 1592 NSTQQCNDAKHRCNOACRAIOETVENKRKERSGGOTNFVLKANVOPEPYK 1643
Result 4	
ID	AAM22482 standard; Protein: 2710 AA. AAM22482
Ac	AAM22482;
Df	07-OCT-1997 (first entry)
Xx	Plasmodium Proj3.
Kw	DBL gene family; SABB; sialic acid binding protein; vaccine; therapy; Duffy binding like gene; Duffy antigen binding protein; erythrocyte; DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response; Plasmodium.
Rv	Plasmodium falciparum.
Xx	
Os	M09640766-A2.
Pn	PN
Pd	19-DEC-1996.
Xx	
Pf	07-JUN-1996; 96WO-USO9508.
Xx	
Pr	07-JUN-1995; 95US-0487826.
Xx	
Pa	(USSH) US DEPT HEALTH & HUMAN SERVICES.
Xx	
Pi	Chitnis C, Miller LH, Peterson DS, Sim KL, Su X; Wellens TE;
Xx	
Dr	WPI: 1997-052231/05.
N-PSDB:	AAT72897.
Xx	

PT	New malaria vaccines - contains cysteine-rich DBL family protein binding domains homologous domains of the Duffy and sialic acid binding proteins
PS	Disclosure; Page 50-56; 96pp; English.
PB	This sequence represents Proj3 of Plasmodium. Proj3 belongs to the Duffy binding like (DBL) family of genes which have homology to the CC CC conserved regions (see AA72889 and AA72888 respectively). The var family of genes modulate cytoadherence and antigenic variation of Plasmodium infected erythrocytes. SAMP and the duffy antigen binding protein (DABP) are soluble proteins that appear in the culture supernatant after infected erythrocytes release merozoites. DABP and SAMP mediate the binding of merozoites and schizonts to the erythrocyte surface. These proteins are necessary for erythrocyte invasion by the parasite. This sequence can be used in the compositions of the invention. The compositions are for the treatment and prevention of malaria, and comprise either a nucleotide sequence or encoded polypeptide of the var-1, var-2, var-3 or var-7 gene of the DBL gene family, a family of genes having homology with conserved regions of DABP and SAMP. The compositions are used for the treatment and prevention of malaria. They are also used in the preparation of vaccines for inducing a protective immune response in a mammal to plasmodium merozoites (especially Plasmodium falciparum or Plasmodium vivax).
SQ	Sequence 2710 AA;
Query Match	Best Local Similarity 40.4%; Score 624; DB 18; Length 2710; Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11
OY	DCHPRKNSNGYPDWC-GNINLVEDPPVCMPPRROKLCVFPLANDNEIKLOSVNLKEA 59 : : : : : : : : : :
Db	1367 ECPNPRES---YDMCKNNIDISH-DGACMPRROKLCLTYIAHESQEFTENITDDNLKA 1422 : : : : : : : : :
OY	60 FTKSAAEFFSWIYYKK-K-DGEWELDKELKGKITPAFLFSLSMETTFEDDYDFDLGTDI 118 : : : : : : : : :
Db	1423 FKTTAAEFTEFLSWMQYKSKNDSEAKILDR----GLTPSQOFLISMWTTFEDYRIDICLTPI 1478 : : : : : : : : :
OY	I19 SKGHGGSLRKOIDISLFENGDSKPNGKTROEMWHESHHEIMVALCALVIKGAKKD- 177 : : : : : : : : : :
Db	1479 SKKDNYAVAKARKICKSFSSDGSKSPSGLSRDGMWMTNGPELMKWGCALTXYVTDTDNK 1538 : : : : : : : : : :
OY	I78 --FTFNKYGNNAKFSDK-STLEFAFKRPQLFRWLIEWDDCYTRQKIADVOEKCS- 233 : : : : : : : : : :
Db	1539 RRIKNIDYSVDKNOONGNQSPGLEEFAKQFLRMWIMEWEFECAEGKOKENITKDNCNI 1598 : : : : : : : : : :
OY	234 NDOLKCDT---ECKNKCEDYVKYM-KKKEWI-----PDQRYYK 268 : : : : : : : : : :
Db	1599 NSTOOCNDAKHRCNOACRAYOEYVENKKREFGQTNFVLKANVGODDEPK 1650 : : : : : : : : : :
RESULT 5	
ID	AA777904
AC	AA777904 standard; Protein; 2710 AA.
XX	AA777904;
DT	13-JUN-2000 (first entry)
XX	
DE	F. falci-parum Proj3 binding domain polypeptide.
KW	DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein; DABP; Sialic Acid Binding Protein; SAMP; malaria; vaccine; Immunisation; Protozocicide; Proj3:
OS	Plasmodium falciparum.
XN	US5993827-A.
PN	
PD	30-NOV-1999.

PF 07-JUN-1995; 95US-0487826.
XX
PR 10-SEP-1993; 93US-0119677.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;
XX WPI; 2000-194198/17.
DR N-PSDB; AAZ98286.
XX
PT Isolated protein binding domains from Plasmodium vivax and Plasmodium
PT falciparum erythrocyte binding proteins useful for vaccinating against
XX malaria -
XX
PS Disclosure; Columns 79-92; 93pp; English.
XX
CC The invention relates to ebl-1 polypeptides that are encoded by the DBL
CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
CC Binding Protein (SABP), which are soluble proteins that appear in the
CC culture supernatant after erythrocytes infected with malaria release
CC merozoites. Immunochemical studies indicate that DABP and SABP are the
CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
CC used to vaccinate against malaria, especially caused by P. falciparum.
CC Immunization with the polypeptide provides effective protection against
CC malaria. The present sequence represents the Proj3 binding domain
XX polypeptide.
XX
SQ Sequence 2710 AA;
XX
Query Match 40.4%; Score 624; DB 21; Length 2710;
Best Local Similarity 43.8%; Pred. No. 4.6e-50;
Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11;
XX
QY 1 DCHPKNSNGYPMQC-GNINIVEDPRVCPRRQRKLCVHFLANDNEIKKLSQVNLKEA 59
DB 1367 ECHPKNS---YPMDCCKNNIDISHD-GACMPRRQRKLCVYIAHESQETNITDDNLKA 1422
XX
QY 60 FIKSAAEFFFSWYKSK-DGEGNLDLKEGKIPPAFLNSMYTFEDYDFLEGTDI 118
DB 1423 FIKTAAAEFTLSWYKSKNDSEAKILDR---GLIPSOFLSKMYTFEDYDFIDICLNTDI 1478
XX
QY 119 SKHGSGSLKEQIDSLFKNGDQSPNGKTRQEMWTEHSEIWEALCALVYIGAKKDD- 177
DB 1479 SKKQNDVAKAKDKIGKFEKSGSKSPSGLSRQEMWTKTNGPEIKKMLCALVYITDTDK 1538
XX
QY 178 --FTENGYNNVKEFSK-STILEEPAKRPQFLRWLTETWYDDCYTRQKYLKDQEKCKS- 233
DB 1538 RKIKNDYSYDKVNOSONGNSPILSEFPAKRPQFLRWMIEMGEFCAEROKENITKDACNEI 1598
XX
QY 234 NDQKCDT---ECNKKCEDYVYKMK-KKKEWT-----PODKYYK 268
DB 1599 NSTQOCNDAKKHCNOACRAIYQETVENKKKEFSGQTNNEVLKANVOPDPEYK 1650
XX
RESULT 6
AAW22475
ID AAW22475 standard; Protein: 3060 AA.
XX
AC AAW22475;
XX
DT 12-SEP-1997 (first entry)
XX
XX Plasmodium var-7.
XX
XX DBL gene family; SABP: sialic acid binding protein; vaccine; therapy;
XX Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
XX DABP, merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
XX Plasmodium.
XX
OS Plasmodium vivax.

OS Plasmodium falciparum.
XX
PN WO9640766-A2.
XX
PD 19-DEC-1996.
XX
PF 07-JUN-1996; 96WO-US09508.
XX
PR 07-JUN-1995; 95US-0487826.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
XX Wellens TE;
XX
XX WPI; 1997-052231/05.
XX
XX N-PSDB; AAT72882.
XX
XX New malaria vaccines - contains cysteine-rich DBL family protein
XX binding domains homologous domains of the Duffy and sialic acid
XX binding proteins
XX
XX Claim 8; Page 61-67; 96pp; English.

XX
XX This sequence represents var-7 of Plasmodium. Var-7 belongs to
XX the Duffy binding like (DBL) family of genes which have homology to the
XX Duffy antigen binding protein (DABP) and sialic acid binding protein
XX (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
XX var family of genes modulate cytoadherence and antigenic variation of
XX Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
XX supernatant after infected erythrocytes that appear in the culture
XX surface. These proteins are necessary for erythrocyte invasion by the
XX parasite. These proteins can be used in the compositions of the
XX invention. This sequence can be used in the compositions of the
XX malaria, and comprise either a nucleotide sequence or encoded polypeptide
XX of the var-1, var-2, var-3 or var-7 genes of the DBL gene family.
XX The compositions are used for the treatment and prevention of malaria.
XX They are also used in the preparation of vaccines for inducing a
XX protective immune response in a mammal to Plasmodium merozoites
XX (especially Plasmodium falciparum or Plasmodium vivax).
XX
SQ Sequence 3060 AA;
XX

Query Match 40.4%; Score 624; DB 18; Length 3060;
Best Local Similarity 43.8%; Pred. No. 5.4e-50;
Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11;
XX

QY 1 DCHPKNSNGYPMQC-GNINIVEDPRVCPRRQRKLCVHFLANDNEIKKLSQVNLKEA 59
DB 1365 ECHPKNS---YPMDCCKNNIDISHD-GACMPRRQRKLCVYIAHESQETNITDDNLKA 1420
XX
QY 60 FIKSAAEFFFSWYKSK-DGEGNLDLKEGKIPPAFLNSMYTFEDYDFLEGTDI 118
DB 1421 FIKTAAAEFTLSWYKSKNDSEAKILDR---GLIPSOFLSKMYTFEDYDFIDICLNTDI 1476
XX
QY 119 SKHGSGSLKEQIDSLFKNGDQSPNGKTRQEMWTEHSEIWEALCALVYIGAKKDD- 177
DB 1477 SKKQNDVAKAKDKIGKFEKSGSKSPSGLSRQEMWTKTNGPEIKKMLCALVYITDTDK 1536
XX
QY 178 --FTENGYNNVKEFSK-STILEEPAKRPQFLRWLTETWYDDCYTRQKYLKDQEKCKS- 233
DB 1537 RKIKNDYSYDKVNOSONGNSPILSEFPAKRPQFLRWMIEMGEFCAEROKENITKDACNEI 1596
XX
QY 234 NDQKCDT---ECNKKCEDYVYKMK-KKKEWT-----PODKYYK 268
DB 1597 NSTQOCNDAKKHCNOACRAIYQETVENKKKEFSGQTNNEVLKANVOPDPEYK 1648
XX

RESULT 7
AAAT77905

ID AAY77905 standard; protein; 3060 AA.
XX
AC AAY77905;
XX
DT 13-JUN-2000 (first entry)
XX
DE Plasmodium var-7 polypeptide.
XX
KM DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KW DABP; Sialic Acid Binding Protein; SAMP; malaria; vaccine; immunisation;
XX protozoacide; var-7.
XX Plasmodium sp.
XX US5993827-A.
PN
PD 30-NOV-1999.
XX
PF 07-JUN-1995; 95US-0487826.
XX
XX 10-SEP-1993; 93US-0119677.
XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Slim KL, Chittis C, Peterson DS, Su X, Wellens TE, Miller LH;

DR WPI; 2000-194198/17.

XX N-PSDB; AA598287.

PT Isolated protein binding domains from Plasmodium vivax and Plasmodium
falciparum erythrocyte binding proteins useful for vaccinating against
malaria -

XX Disclosure; Columns 109-124; 93pp; English.

XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
CC Binding Protein (SABP), which are soluble proteins that appear in the
CC culture supernatant after erythrocytes infected with malaria release
CC merozoites. Immunochemical studies indicate that DABP and SABP are the
CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
CC used to vaccinate against malaria, especially caused by P. falciparum.
CC Immunization with the polypeptide provides effective protection against
CC malaria. The present sequence represents the var-7 polypeptide.
XX

SO Sequence 3060 AA;

Query Match 40.4%; Score 624; DB 21; Length 3060;

Best Local Similarity 43.8%; Pred. No. 5.4e-50;

Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11;

QY 1 DCHPRKNSNGYPMOC-GNINLVEDPRVCMPPRRQKLCYHFLANDNEIKKLSQVNLKEA 59
DB 1365 ECHPKS---YPMDCNNIDISHD-GACMPPRQKLCYHFLANDNEIKKLSQVNLKEA 1420
QY 60 FIKSAAEETFFSWYYKSK-DGEGNELDKELKGIKIPPAFLRSMFYTFGGYRDLFGTDI 118
DB 1421 FIKTAAEFTLSWYYKSKNDSEAKILDR---GLIPSGFLRSMFYTFGGYRDLFGTDI 1476
QY 119 SKHGSGSKLEKIDSLFRKNGDQSPNGKTRQEMWTEHSHETIEMALCALVTKGAKKD- 177
DB 1477 SKRQNVAAKAKDKIGFFSKDGSKSPSGLSRQSMWMTNGEPIKMGMLCALTKYVTDNDK 1536
QY 178 --FTENYGVNNVKSFK-STTLEEFARPOFLRMLTEWYDLCYTRQYKLDQOEKCS- 233
DB 1537 RRTKNDYSIDKVNQSONGNFSLSEFAKPOFLRMLTEWYDLCYTRQYKLDQOEKCS- 1596
QY 234 NDLQKDTF---EKNKCEDYVYKMK-KKKEMI-----PDKYK 268
DB 1597 NSTQCCNDAKHRCNQACRAVOEYENKKEFFSGOTNNFVLKAVNQPDPEYK 1648

RESULT 8
ID AAB62148
XX AAB62148 standard; peptide; 407 AA.
XX

AC AAB62148;

DT 29-MAY-2001 (first entry)

DE P. falciparum varCSA polypeptide A4 DBL4-gamma.

XX FCRI3, varCSA protein; chondroitin sulfate A (CSA); var gene; PFEMP1;
KW erythrocyte membrane protein 1; parasitized red blood cell; PIRC;
XX malaria; protozoacide; A4 DBL4-gamma.
XX Plasmodium falciparum.

OS WO200116326-A2.

PN 08-MAR-2001.

PD 01-SEP-2000; 2000WO-US24195.

PF 01-SEP-1999; 99US-0152023.

PR (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;

XX Gysin J, Pouvelle B, Fujil N, Smith J;

DR WPI; 2001-235109/24.

XX Novel FCRI3, varCSA protein, useful for modulating parasitized red blood
PT cell binding, sequestration and onset of maternal malaria -
XX

PS Disclosure; Page 72-73; 78pp; English.

XX The invention relates to a P. falciparum FCRI3, varCSA protein, that is
CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
CC corresponding P. falciparum erythrocyte membrane protein 1 (PFEMP1)
CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
CC protein and the encoding gene are useful for treating and preventing
CC maternal malaria in a patient afflicted with a risk for contracting
CC maternal malaria or in a patient afflicted with maternal malaria. The
CC present sequence represents a P. falciparum varCSA polypeptide
XX A4 DBL4-gamma.

SO Sequence 407 AA;

Query Match 39.5%; Score 609; DB 22; Length 407;

Best Local Similarity 42.3%; Pred. No. 1.1e-49;

Matches 127; Conservative 50; Mismatches 89; Indels 34; Gaps 10;

QY 1 DCHPRK--KNSNGYPMOCGNINLVEDPRVCMPPRRQKLCYHFLANDNEIKKLSQVNLKE 58
DB 11 ECHPKYIPYNDPGWNTCKVINREKSGCMPPRRQKLCI---NEHLEKATETELRK 67
QY 59 AFIKSAAEETFFSWYYK-----SKDEG-----NELDKELKGIKIPPAFLRSMFYTF 106
DB 68 AFIECAALIEFWLMDKKEKDEKKEGGISDPPDPKLEGGITIPEDFRKQNTTY 127
QY 107 GDYRDLFGTDISKHGSGSKLEQIDSLFRKNGDQSPNGKTRQEMWTEHSHETIEMALC 166
DB 128 GDYRDLFGTDISKHGSGSKLEQIDSLFRKNGDQSPNGKTRQEMWTEHSHETIEMALC 187
QY 167 AL---VRIKAKDFTFNIVYNNVKSFSKSTTLEEFARPOFLRMLTEWYDLCYTRQYK 223
DB 188 GLSHHILNK--NKQRLRNLDNN--KYTKISSKLEDFASRQFLRMLTEWYDLCYTRQYK 245
QY 224 LKDVQEKSGNDQKDTPE-----CNKKECEDYVYKMKKKKEMIPOD--KYKDEBDK 274
DB 246 INDLKTC---NVEEGSGNGKREKCKNACENAVKSLKMDKMQYEQGTAKFKDKDKK 302

260 NKTCDNENCGACKTQCEKYYKKMERWK-----KHYSSO--KKKF 297

[illegible]

RESULT 11
 ID AAR70234 standard; Protein; 793 AA.
 XX AAR70234;
 AC AAR70234;
 XX 22-SEP-1995 (first entry)
 XX P. falciparum E31a.
 DE P. falciparum E31a.
 XX Erythrocyte binding ligand; E31a; binding domain; malaria; therapy;
 KM vaccine.
 XX Plasmodium falciparum.
 OS Plasmodium falciparum.
 XX WO9507353-A.
 PN 16-MAR-1995.
 XX 07-SEP-1994; 94WO-US10230.
 PR 10-SEP-1993; 93US-0119677.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Chitnis C, Miller LH, Peterson DS, Slim KL, Su X;
 PI Wellens TE;
 XX WPI: 1995-123427/16.
 DR N-PSDB; AAO83527.
 XX New erythrocyte binding domain polypeptide(s) - isolated from
 PT Plasmodium binding proteins, used in diagnosis, treatment and
 PR prevention of malaria
 XX Disclosure: Page 51-52; 81pp; English.

XX Erythrocyte binding ligand (EBL) family genes were cloned from
 CC P. falciparum chromosome 7 subsegment libraries constructed during
 CC genetic studies of the chloroquine resistance locus. The 4 genes,
 CC EBL-e1 (AA083526), E31a (AA083527), EBL-e2 (AA083528) and Pro3
 CC (AA083529), encode the proteins given in AAR70233-36, respectively. The
 CC binding domains of such proteins can be expressed e.g. in E. coli,
 CC yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
 CC cells, and provide protection against P. falciparum.

SO Sequence 793 AA:

Query Match 32.0%; Score 493.5; DB 16; Length 793;
 Best Local Similarity 40.0%; Pred. No. 2,8e-38;
 Matches 124; Conservative 39; Mismatches 94; Indels 53; Gaps 15;

QY 1 DCHPRKNSGYDPMOCGNL-VEPRVCMPPRROKLCVHFLANDNEIKLOSQV-LKE 58
 DB 414 EYRKET--YSEWTCDESKIKMGHGACIPPRKOLCHYL-----EKIMTNTELKY 464
 QY 59 AFIKSAATETFSWY-YRKSDEGENELDELKEKIPPAFLKSMFYTFGQYNDFFLFGTD 117
 DB 465 AFIKCAAAETFLMOMYKKDMGNAEDDELKLGITIPDEFKQMYTTADYDIDLGTD 524
 QY 118 I-----SKGHEGSKLQKQISLFRNGDQKSPNG-KTQOEWTEHSHEIWEAMCAL-- 168
 DB 525 ISSKRDTSKGVG---KVKCNIDVF---YKISNSIRYKRSWETNGPVIWEGMICALSY 577
 QY 169 -----VKIGAKKDPTEFYGVNNVKE-SDKSTTLEEFAPKPOELRWLEWYDYDCTQ 221
 DB 578 DTSINAVNETAKKLTGNNNFEKVTIFGSDSTTLKFSERPOFLRWLEWGENEFCRQ 637
 QY 222 KYLDVQERCKNSDQKCDTECNK---CEDYVK-----YMKKKKMIQDQ- 265
 DB 638 KEKVKVLACKDCD-VDGQKCKGKCVACKDQCKOYHSMIGIWDYKKQKGYTEVKKI 696

QY 266 -YKDERDKK 274
 DB 697 PLYKEDKDYK 706

RESULT 12
 ID AAM22480 standard; Protein; 921 AA.
 XX AAM22480;
 AC AAM22480;
 XX 07-OCT-1997 (first entry)
 XX Plasmodium E31a.
 DE Plasmodium E31a.
 XX DBL gene family; SABB; stailic acid binding protein; vaccine; therapy;
 KM Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KM DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KM Plasmodium.
 XX Plasmodium falciparum.
 OS Plasmodium falciparum.
 XX WO9640766-A2.
 PN 19-DEC-1996.
 PD 07-JUN-1996; 96WO-US09508.
 PF 07-JUN-1995; 95US-0487826.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Chitnis C, Miller LH, Peterson DS, Slim KL, Su X;
 PI Wellens TE;
 XX WPI: 1997-052231/05.
 DR N-PSDB; AAT72895.

XX New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and stailic acid
 PR binding proteins
 XX Disclosure: Page 43-45; 96pp; English.

XX This sequence represents E31a of Plasmodium. E31a belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and stailic acid binding protein
 CC (SABB) conserved regions (see AAT72889 and AAT72888 respectively). The
 CC var family of genes modulate cytoadherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABB and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and
 CC SABB mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the
 CC invention. The compositions are for the treatment and prevention of
 CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
 CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family. a
 CC family of genes having homology with conserved regions of DABP and SABB.
 CC They are also used in the preparation of vaccines for inducing a
 CC protective immune response in a mammal to Plasmodium merozoites
 CC (especially Plasmodium falciparum or Plasmodium vivax).

SO Sequence 921 AA:

Query Match 32.0%; Score 493.5; DB 18; Length 921;
 Best Local Similarity 40.0%; Pred. No. 3.4e-38;
 Matches 124; Conservative 39; Mismatches 94; Indels 53; Gaps 15;

QY 1 DCHPRKNSGYDPMOCGNL-VEPRVCMPPRROKLCVHFLANDNEIKLOSQV-LKE 58
 DB 414 EYRKET--YSEWTCDESKIKMGHGACIPPRKOLCHYL-----EKIMTNTELKY 464

QY 59 AFKSAAEFFSWY-YRKSNDGEGNELDKELKEKIPAPLRSMTFFGDRFLGTD 117
 DB 465 AFKCAAAETFLMKNYKDKNGNEDDEKLGKGIIPEDRKQMTFFADYRICIGTD 524
 QY 118 I-----SKGHEGSKLEQIDSLFKNGDQSPNG-KTROEMWTEHSHIEWAMICAL-- 168
 DB 525 ISSKSDTSKGVG--KVCNCIDDFV---YKISNSIRYRKSMWETNGPVWEGMICALSY 577
 QY 169 -----VKIGAKKDDFTENYGYNNVKF--SDKSTLEEARAKPOPLRWLTGYDYCYTRQ 221
 DB 578 DTSLNANVPETHKKLTBEGNNNEFKVITGSDSTLTSKFSERPQPLRWLTGEGNEPCKEOK 637
 QY 222 KYLDVQCKCKSNDQKCDTECNK---CEDYK-----YMKKKKEWIPQDK- 265
 DB 638 KEYVVLAKCKDCD-VDGDKCNGKCVACQCKQYHSMWIGWIDNKKQKGRTEYVKI 696
 QY 266 -YYKDERDK 274
 DB 697 PLYKEDKDVK 706

RESULT 13
 AAY77902
 ID AAY77902 standard; Protein; 921 AA.
 AC AAY77902;
 DT 13-JUN-2000 (first entry)
 DE P. falciparum ebl-1 related polypeptide.
 KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
 KM DABP; Sialic Acid Binding Protein; SAMP; malaria; vaccine; Immunisation;
 XX Protozoacide.
 OS Plasmodium falciparum.
 XX US5993827-A.
 XX 30-NOV-1999.
 PD 07-JUN-1995; 95US-0487826.
 PF 10-SEP-1993; 93US-0119677.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Slim KL, Chienis C, Peterson DS, Su X, Wellens TE, Miller LH;
 DR MPI: 2000-194198/17.
 N-PSDB; AA298284.

PT Isolated protein binding domains from Plasmodium vivax and Plasmodium
 PT falciparum erythrocyte binding proteins useful for vaccinating against
 PT malaria -
 XX
 PS Disclosure: Columns 61-66; 93pp; English.
 XX
 CC The invention relates to ebl-1 polypeptides that are encoded by the DBL
 CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
 CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
 CC Binding Protein (SABP), which are soluble proteins that appear in the
 CC culture supernatant after erythrocytes infected with malaria release
 CC merozoites. Immunochemical studies indicate that DABP and SABP are the
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
 CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
 CC used to vaccinate against malaria, especially caused by P. falciparum.
 CC Immunization with the polypeptide provides effective protection against
 CC malaria.
 CC
 XX Sequence 921 AA;

Query Match 32.0%; Score 493.5; DB 21; Length 921;
 Best Local Similarity 40.0%; Pred. No. 3.4e-38;
 Matches 124; Conservative 39; Mismatches 94; Indels 53; Gaps 15;

QY 1 DCHPRKNSNGYIPDMOGNINL-VEDPVCMPPRRKLCVHFLANDNEIKLQSOVN-LKE 58
 DB 414 ECTRKET---YSEWTCDESKIKMGHACIPPRKRLCIHYL-----EKIMTNTNELKY 464

QY 59 AFKSAAEFFSWY-YRKSNDGEGNELDKELKEKIPAPLRSMTFFGDRFLGTD 117
 DB 465 AFKCAAAETFLMKNYKDKNGNEDDEKLGKGIIPEDRKQMTFFADYRICIGTD 524
 QY 118 I-----SKGHEGSKLEQIDSLFKNGDQSPNG-KTROEMWTEHSHIEWAMICAL-- 168
 DB 525 ISSKSDTSKGVG--KVCNCIDDFV---YKISNSIRYRKSMWETNGPVWEGMICALSY 577
 QY 169 -----VKIGAKKDDFTENYGYNNVKF--SDKSTLEEARAKPOPLRWLTGYDYCYTRQ 221
 DB 578 DTSLNANVPETHKKLTBEGNNNEFKVITGSDSTLTSKFSERPQPLRWLTGEGNEPCKEOK 637
 QY 222 KYLDVQCKCKSNDQKCDTECNK---CEDYK-----YMKKKKEWIPQDK- 265
 DB 638 KEYVVLAKCKDCD-VDGDKCNGKCVACQCKQYHSMWIGWIDNKKQKGRTEYVKI 696
 QY 266 -YYKDERDK 274
 DB 697 PLYKEDKDVK 706

RESULT 14
 AAB62150
 ID AAB62150 standard; peptide; 311 AA.
 AC AAB62150;
 DT 29-MAY-2001 (first entry)
 DE P. falciparum varCSA polypeptide FC3 var3DBL-gamma.
 XX
 KW FC3 varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
 KM erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
 XX malaria; protozoacide; FC3 var3DBL-gamma.
 OS Plasmodium falciparum.
 XX WO200116326-A2.
 XX 08-MAR-2001.
 PD 01-SEP-2000; 2000WO-US24195.
 PF 01-SEP-1999; 99US-0152023.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
 PI Gysin J, Pouvelle B, Fujii N, Smith J;
 DR MPI: 2001-235109/24.
 XX
 PT Novel FC3 varCSA protein, useful for modulating parasitized red blood
 PT cell binding, sequestration and onset of maternal malaria -
 XX
 PS Disclosure: Page 74-75; 78pp; English.
 XX
 CC The invention relates to a P. falciparum FC3 varCSA protein, that is
 CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
 CC corresponding P. falciparum erythrocyte membrane protein 1 (PFEMP1)
 CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
 CC protein and the encoding gene are useful for treating and preventing
 CC maternal malaria in a patient afflicted at a risk for contracting
 CC maternal malaria or in a patient afflicted with maternal malaria. The
 CC present sequence represents a P. falciparum varCSA polypeptide

CC FCR3 var3DBL-gamma.
 XX
 SQ Sequence 311 AA;
 Query Match 31.5%; Score 486; DB 22; Length 311;
 Best Local Similarity 39.1%; Pred. No. 4.3e-38;
 Matches 115; Conservative 39; Mismatches 108; Indels 32; Gaps 13;

QY 5 KNSNGYPMOCGMINIYEDPRVCMPPRRQKLCVHFLANDNEIKLQSQVNLKEAFIKSA 64
 DB 13 KAGKDKYPMDCNSQIHTHNGACMPRRQKLCVSLGTDRKALE--YIRTEIKSA 69
 QY 65 AAEFTFSWYIYKSGOGEGLDEKELKEGKIPPAFLRSMFTYTGDIYDFLFGDISGQHB 124
 DB 70 ALETHAMORIKEDNGEA---EAEIKNGNIPEFGFKROMYITTEGDIYDFLFGDISGQHB 125
 QY 125 GSKLKEOISLF-KINGDOK-----SPNGKROEMWTEHSEIWEAMCALV-KIG--AKK 175
 DB 126 ISGVSEKVIITILEKENDAKYAAKONSNNELDDMDMDQHGKDIWEGMICALTKISDEBK 185
 QY 176 DDEFTENYNNVAFSDK-STILEEFAKRPQFLMTEYTDYCYTRQKYLKDYQEC-- 231
 DB 186 KEIKNNYSYKTLNESPKGSKNKEDEFAKQPLFMFTIEMGDEFCAGREKEAKYKVCSDA 245
 QY 232 KSDQKLCOTECNKKC-----EDYVKYMKKKKEWIPQDKYKDER--DKKRF 276
 DB 246 KDYDGCK-WTKSNASCVSACKVYEDYI--TKKVEYTKQKGFDAEKIDKEGY 296

RESULT 15
 AAW93944
 ID AAW93944 standard; protein; 2228 AA.
 AC AAW93944;
 XX
 DT 30-JUN-1999 (first entry)
 XX
 DE P. falciparum PfEMP1 protein.
 XX
 KW Erythrocyte membrane protein; EMP; PfEMP1; malaria; antioocclusion;
 KW glycosaminoglycan-like moiety; antiaggregation; antimalarial;
 KW antigen receptor; infected erythrocyte; rosette formation; blood cell;
 KW capillary occlusion; cerebral malaria; treatment; vaccine; detection;
 KW medication; parasite; diagnosis; drug screening.
 OS Plasmodium falciparum.
 XX
 WO9915557-A1.
 01-APR-1999.
 PF 18-SEP-1998; 98WO-SE01675.
 XX
 PR 19-SEP-1997; 97SE-0003386.
 XX
 PA (KARO-) KAROLINSKA INNOVATIONS AB.
 XX
 PI Barragan A, Carlson J, Fernandez V, Oljun C, Wahlgren M;
 XX
 DR WPI; 1999-254692/21.
 XX
 PS New isolated malaria polypeptides
 PT
 XX
 PS Claim 4; Page 67-74; 80pp; English.
 CC This invention describes a novel Plasmodium falciparum erythrocyte
 CC membrane protein (EMP), PfEMP1, which is capable of binding to a
 CC carbohydrate which exhibits at least one negatively charged
 CC glycosaminoglycan (GAG)-like moiety and has antiaggregational,
 CC antioocclusional and antimalarial activity. The carbohydrates of the
 CC invention are capable of acting as receptors for malaria antigens
 CC present on the surfaces of malaria infected erythrocytes, by binding
 CC to these antigens the carbohydrates prevent rosette formation by the

CC blood cells, this prevents occlusion of capillaries as is seen in
 CC cerebral malaria caused by Plasmodium falciparum. The products of the
 CC invention can be used to treat malaria or to vaccinate against it, or
 CC used to design a model to identify compounds that bind to PfEMP1. The
 CC carbohydrates, polypeptides and antibodies of the invention can be used
 CC as a medication for dissolving the rosettes formed by erythrocytes
 CC infected by a malaria parasite. The products can also be used for
 CC detection, diagnosis and drug screening.

QY 8 SNGYPMOC-----GNINLEDPR--VCMPPRRQKLCVHFL-----ANDNEIKL 50
 DB 911 NNSRLGRCVTPSGEFTTSQKGAICVPRRRRLYIKKIYVMATYTESPQASGSASST 970
 QY 51 -----QSQVNLKEAFIKSAAEFTFSWYIYK-----SKDEGNEL----- 85
 DB 971 SGSTPPDSKELKAFVESAAIETFFLMHRYKEEKKAVALQEGAGHGLPRVEEGSPEDP 1030
 QY 86 DKELKSGKIPPAFLRSMFTYTGDIYDFLFG--IDISK-----GQHB 124
 DB 1031 EDKLEKGIPLDGLRQMFYTLGDIYDFLFGSNDTTSVSKDPPSSNDNLKNIYLLASGS 1090
 QY 125 GSKLKEOISLF--FKN-----GDQSPNGKTR--QEWTEHSEIWEAMCALV--KIG- 172
 DB 1091 TEOERKMKYKIKIKFRKRCSTERSAPNLVSHQYTWENNNGKTIYHGMICALSKIKIK 1150
 QY 173 --AKKDDFTEN-----YGYNNVAFSDKS-----TTLE 197
 DB 1151 GVEKKRPQKIENPENLWDEANKKRPQYQYTYNKLDENSGTSPTTQTOASSDNTFTTL 1210
 QY 198 EFAKRPQFLMTEYTDYCYTRQKYLKDYQEC--SNDQKCT----- 241
 DB 1211 HFVKRPYFRWFEWGESFCREKRRKRLKQIKVCKVENDVGRSGSDGACDSISTHDYS 1270
 QY 242 -----EONKCEDYVYMKKK--EWIPQDKYKDER 271
 DB 1271 TVPSFNCPCGCKHSCSSYRKWIEKKIEFHKSANAYQOK 1309

Search completed: April 28, 2003, 10:30:10
 Job time : 14.4199 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 5.10497 Seconds

(without alignments)
5197.501 Million cell updates/sec

Title: US-10-087-013-2-COPY_1279_1554

Perfect score: 1543
Sequence: 1 DCHPRKNSNGTPDMQGNIN.....KKEWIPDQKYKDERDKRRF 276

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

tal number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	634.5	41.1	2706	2 T28155	variant-specific s
2	624	40.4	3078	2 T28432	variant-specific s
3	609	39.5	3026	2 T28431	variant surface pr
4	486	31.5	3006	2 T28625	variant-specific s
5	475	30.8	1711	2 C71625	variant-specific s
6	410.5	26.6	2042	2 T18399	variant-specific s
7	402.5	26.1	2135	2 T14602	variant-specific s
8	396	25.7	2212	2 T28157	variant-specific s
9	390.5	25.3	2228	2 T14029	erythrocyte membra
10	385.5	25.0	2664	2 T28626	variant-specific s
11	382.5	24.8	2182	2 T28634	variant-specific s
12	371.5	24.1	2647	2 T28161	variant-specific s
13	369.5	23.9	2924	2 T18378	variant-specific s
14	339.5	22.0	2197	2 B1600	variant-specific s
15	299.5	19.4	1729	2 T18396	erythrocyte membra
16	274	17.8	1435	2 A37793	erythrocyte-bindin
17	236	15.3	1153	2 T28652	erythrocyte bindin
18	223.5	14.5	1045	2 T18373	erythrocyte bindin
19	223	14.5	1070	2 T30848	Duffy receptor - p
20	172.5	11.2	778	2 A35970	erythrocyte-bindin
21	106	6.9	3119	2 T18414	protein g377 - mal
22	105.5	6.8	346	2 D64484	hypothetical prote
23	104.5	6.8	378	2 T18486	hypothetical prote
24	104.5	6.8	781	2 F90547	hypothetical prote
25	103.5	6.7	677	2 B29599	DNA-directed RNA p
26	101	6.5	522	2 C96608	hypothetical prote
27	100.5	6.5	980	2 E71606	hypothetical prote
28	100	6.5	227	2 D90512	conserved hypothet
29	99.5	6.4	759	2 D71853	hypothetical prote

30	99	6.4	1650	2 T18444	hypothetical prote
31	99	6.4	3660	1 S02041	dystrophin, muscle
32	98.5	6.4	1345	2 A64430	DNA-directed RNA p
33	98.5	6.4	1467	2 T23950	hypothetical prote
34	98	6.4	2010	2 B71616	phosphatase (acid
35	97	6.3	963	2 C90535	conserved hypothet
36	96.5	6.3	938	2 G70472	hypothetical prote
37	95.5	6.2	678	2 A54514	glutamic acid-rich
38	95	6.2	1391	2 T20642	hypothetical prote
39	95	6.2	1397	2 E87998	protein F09C3.111
40	95	6.2	4427	2 PNO637	polyketide synthas
41	94.5	6.1	446	2 A48407	neutral class-II be
42	94.5	6.1	1178	2 S78475	mannosylphosphoryl
43	94	6.1	1081	1 A42399	isolectine-tRNA 11
44	93.5	6.1	612	2 C72293	hypothetical prote
45	93	6.0	428	2 A45568	highly immunoreact

ALIGNMENTS

RESULT 1

T28155

variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum) (fragment)

N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)

C:Species: Plasmodium falciparum

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C:Accession: T28155

R:Rowe, J.A.; Moulds, J.M.; Newbold, C.I.; Miller, L.H.

Nature 388, 292-295, 1997

A:Title: Plasmodium falciparum reosetting is mediated by PfEMP1 and requires complemen

A:Reference number: Z20477; PMID:97373957; PMID:9230440

A:Accession: T28155

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2706 <ROW>

A:Cross-references: EMBL:Y13402; PIDN:CAA7831.1

A:Experimental source: strain IT 4/25/5

C:Genetics:

A:Introns: 2493/3

A:Note: R29R+var1

Query Match 41.1%; Score 634.5; DB 2; Length 2706;
Best Local Similarity 45.5%; Pred. No. 1.2e-39;
Matches 131; Conservative 37; Mismatches 91; Indels 29; Gaps 10;

QY	1	DCHPRKNSNGTPDMQGNINLVE-DPRVCMPPRRQKCYHFLANDNIRKLOQVYNKEA 59
DB	787	NCNAKNRKN-EMQCDKNTFVGNEGVCMPPRRKSICINLTLEQTK--NKYOLREA 841
QY	60	FIRSAAEFFPSWYYSKSGEGNELDKELKEGKIPAPFRRSMFYFGDRDFEGDIS 119
DB	842	FIRCAKENTMLMDKYNKNDNEAEL---LAKGKIPDFRIRHFTYTGDRDCLNDMG 898
QY	120	KHGEESKLEQIDSLFKNDQKSPNGKTRQEWTEHSHIWEIMCALVYKAGK---- 174
DB	899	K--DVDKVKKNINKVFNNSKRGFKKIDPENWNNENGPOIMNGMICALIHADTKDSIKN 955
QY	175	KDQFTEYGYNNVNFSDK-----STLEEFARKPQPLRMLEWYDDVCYRQYIKQVQE 229
DB	956	KD---NYKTEKVTIILAKRDSNGMTLSEFAKPKPFLRMVEWDDCKRQKRYLTVAS 1011
QY	230	KCKSND--QLKCDTECNKCCEDYVYKMKKK-EMIPDQKYKDERDOK 274
DB	1012	TCKSIDQGLKCDRCGNCKNCKDEYKRYKRRKKKEEMNLQDKYKQKREMK 1059

RESULT 2

T28432

variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)

N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)

C:Species: Plasmodium falciparum

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

2122 GYRDFLEGTIDSKGKRESALGKRTIDSLFKNDQKSSGKPTLEWMDYGPDIWKGMVC 2181

1. Molecule type: DNA
 2. Residues: 1-1711 <GAR>
 3. Cross-references: GB:AE001366; NID:g3845070; PID:g384
 4. Experimental source: clone 3D7

RESULT 5
 C71625
 Variant-specific surface protein 1 homolog PFB0010w - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 A:Accession: C71625
 R:Gardner, M.J.; Jettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.; Perle, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; MUID:99021743; PMID:9804551
 A:Accession: C71625
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown.
 A:Molecule type: DNA
 A:Residues: 1-1711 <GAP>
 A:Cross-references: GB:AE001365; GB:AE001362; NID:g3845070; PIDN:AACT1792.1; PID:g384
 A:Experimental source: clone 307
 C:Genetics:
 A:Gene: PFB0010w

Query Match 25.7%; Score 396; DB 2; Length 2212;
Best Local Similarity 27.5%; Pred. No. 1e-21;
Matches 101; Conservative 50; Mismatches 94; Indels 122; Gaps 12;

27 VCPPRRLCV---HFLANDEIKKLSQV-----NLKEAFIKSAAE 67
1342 ICIPRRRLRYOKLHDMASGTVVSGAOTPOGSSPSGKNEPSSDLRAFIQSAIE 1401
68 TFFSWYYSKSDGNEMLDKELK-----GKIP 95
1402 TFFLMDRYKE---KEIEKKKKYANGVPSLNGCPPOPGVYTGDSPOSKLOOTGYIP 1457
96 PAFLSMYTTFDGYRDLFGD---ISKHGE-----GSKLKQIDSLFRNGDOKSPNG 146
1458 PEFLLQMYTTFDGYRDLFGD---ISKHGE-----GSKLKQIDSLFRNGDOKSPNG 1517
147 KTRQEMWTEHSHIEWEAMLCALV-----KIGAK-----KDDFTENTGVN 185
1518 EKROTWMQNGEHIWNGICALTYKREDEKTPKQNGELKSALMDKNNKPKDOKYQD 1577
186 NVKFSKDS-----TLEEFKAPQOFLMTEWYDYCYTRQKYLKDOEKCS 233
1578 KVLIDENSISPKTNDHYPPPLTNFISRPYFRYLEMGETCEREKKRLKIKYECMD 1637
234 ND--OLKCD-----TECNKCEYVYMKKK-EMIPDOKYK 268
1638 EDGKKKCSGDGECEIRKODYSTVADFYCPCGKYCRPKYMKIEKKDEYDOKQEAVN 1697
269 DERKKR 275
1698 NQKTAR 1704

RESULT 9

T14029
Variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
N:Alternate names: erythrocyte membrane binding protein 1 (EMPI)
C:Species: Plasmodium falciparum
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
C:Accession: T14029
R:Chen, Q.; Barragan, A.; Fernandez, V.; Sundstrom, A.; Schlichtherle, M.; Sahlen, A.; J. Exp. Med. 187, 15-23, 1998
A:Title: Identification of Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1)
A:Reference number: Z17860; MUID:96080592; PMID:9419207
A:Accession: T14029
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2228 <CHE>
A:Cross-references: EMBL:AF003473; NID:92961467; PID:92961468; PIDN:AAC05730.1
A:Note: FCRS1.2-vari

Query Match 25.3%; Score 390.5; DB 2; Length 2228;
Best Local Similarity 28.3%; Pred. No. 2.7e-21;
Matches 113; Conservative 48; Mismatches 103; Indels 135; Gaps 18;

8 SNGYDMOC-----GIVNVEPR--VCPPRRLCVHFL-----ANDNEIKKL 50
911 NNSRLGMRCTPSPSEPTSSDKNGAICVPPRRRLYIKIYDNTKTESPOASGSSEASST 970
51 -----SOYNLKEAFIKSAAEFFSWYK-----SKDEGNEL-----85
971 SGTTPPSKALAKAVESAIETFFLMHRYKREKKAIVAOEGAGHGLPVEBGSPEYD 1030
86 DKEIKEKIPPAFLRSFYFGDYRDLFG-----TDISK-----GHGE 124
1031 EDKLEKGIIDGLKQMFYLDGYRDLIFSGNDTISVSDTSSSDNKNIVLASGS 1090
125 GSKLEQIDSL--FKN-----GDOKSPNGKTR-QEWMTESHEIWEAMLCALV--KIG- 172
1091 TEGEREMNNKYEIKNFKRCKSTERSAPNLVSHPTWMENNGKTYIWHGAVCALTSKDKIAK 1150

QY 173 --AKKDFTEN-----YGVNNVFEKSDS-----TILE 197
DB 1151 GVEKKPKIKINPENLMDANKKRPPOYOTVTKIDENSIGTSBRTQOASSDNPTTIT 1210
QY 198 EPAPKROFLMTEWYDYCYTRQKYLKDOEKCK-----SNDLAKCT- 241
DB 1211 HFVRRPYTFEWMESGCEKREKRLKQIKYDCKENKDVRCGSDGACDSTHDS 1270
QY 242 -----BCNKKCEYVYMKKK-EMIPDOKYKDER 271
DB 1271 TVSPFNCPCGKHCSSYRKWKIEFKIEFKOSNAYGOQR 1309

RESULT 10

T28626
Variant-specific surface protein 2 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28626
R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson, Cell 82, 89-100, 1995
A:Title: The large diverse gene family var encodes proteins involved in cytoadherence
A:Reference number: Z20487; MUID:95330813; PMID:7606788
A:Accession: T28626
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2664 <SUX>
A:Cross-references: EMBL:L40609; NID:9886376; PID:9886378; PIDN:AAA75398.1
A:Genetics:
A:introns: 2197/3
A:Note: var-2

Query Match 25.0%; Score 385.5; DB 2; Length 2664;
Best Local Similarity 27.4%; Pred. No. 7.9e-21;
Matches 118; Conservative 44; Mismatches 107; Indels 161; Gaps 17;

QY 2 CHPKNSNGY--PDQCGNINLV-----EDPRVCPPRRLCVHFL--ANDN 45
DB 1313 CIPKTSNDVYTGEGCG-GSRVYRSTPSGSSNDKNGACIPRRRLYVGLQWANKH 1371
QY 46 EIKLQSQ-----VNLEAFIKSAAEFFSWYK- 76
DB 1372 NTFYSVQGEATARSSEAPAPGKSSSGKREPSDKLRTAFIESAAVETFLMDRIK 1431
QY 77 -----SKOEGNELKELKEKIPPAFLRSFYFGDYRDLFG--G 115
DB 1432 EWLAKKAELONGLDLSSGDDPDNPQNKLLNGVLPDLRLMFYTLGDRDLVHCG 1491
QY 116 TDSKGHGBS-----KLEKQIDSLF-KNGDQ--KSPNGKTRQEWMT 154
DB 1492 NTSDSGNTGSSNNNIVLEASGNKEEDMKIQEKIQILPKNGTGLVYKSSAQTPDKMWN 1551
QY 155 EHSHEIWEAMLCALV-----KIGAKKD-----177
DB 1552 EHAESIWKMICALYTYTEKNPDTSGARGDENKI--EKDEVEYEKEFGSTADKHGTASTPTG 1609
QY 178 -FTENYGINNVKFSKDS-----TLEEFKAPQOFLMTEWYDYCYTRQKYLK 225
DB 1610 TYKTOYDEKXVLEPTSAKPPASASDPLSLDYLRLRPYFRYLEMGONCKEKKRLK 1669
QY 226 DVOEKK-----SNDLKCD-----TECNKCEYVYMKK-KEKEWIPQ 263
DB 1670 QIKKECMGSDKRTSGDGEQDRDRTNSVSADEGRSGCNSCFYKWKIKRKRKEDQ 1729
QY 264 DKYYDERDK 273
DB 1730 ANAYSKOKTK 1739

RESULT 11

T28634
Variant-specific surface protein 7 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum

Db 1154 LDDTSANTGSPVSGEKIPPLDIFISRPYFRYLEMGEFTCKERRKRLKIEKIECRG 1213
QY 234 -----NQD---LKCDTECNKCKEDYVYKMK-KKKEMIPQDKY 267
Db 1214 DRTGHEHSGGCGYCTRTDADNRDKFVNLNC-RDCHICCKRTKRWIDIKFDEYHKEKKY 1272
QY 268 KDERDK 273
Db 1273 QGEYDK 1278

RESULT 14

B71600
variant-specific surface protein 1 homolog PFBI055C - malaria parasite (Plasmodium falciparum)
N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: B71600
R:Gardner, M.V.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; Porter, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: B71600
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2197 <GAR>
A:Cross-references: GB:AE001434; GB:AE001362; NID:93845341; PIDN:AC71996.1; PID:9384534
C:Genetics:
A:Experimental source: clone 3b7
A:Gene: PFBI055C

Query Match 22.0%; Score 339.5; DB 2; Length 2197;
Best Local Similarity 25.2%; Pred. No. 1.9e-17;
Matches 107; Conservative 50; Mismatches 114; Indels 153; Gaps 17;

QY 2 CHPKNSNG---YPMOC---GNINLV-----EDPRVCMPPRROKICVHFL-- 41
Db 931 CLKXGPGCKERFPWMCVSSGSEKSVATAGSSGATGKSGDKALICVPRRRRLVYGGGLK 990
QY 42 -----AND---NEIKKLSQVNLKEAFKSAAEFPFSAATFESWYYSK 78
Db 991 LTSAGTSSSEPOGSESSRASDVSOQNGGDDTTTSTSLKMFLETALETFFLHRYKKE 1050
QY 79 -----DGEENELDKEL-KEGKIPPAFLRSFYFGDYDFL 113
Db 1051 WEAKKAEIQRNGLLGTGASLNLGGDSNPQTQLQKSGTIPDLRLMFLYLGYYRDL 1110
114 F-GDIDISGHBG-----SKLEQJDSLFKNGDOKSPNG-----KTRQEW 152
1111 VRGVAADNRKNGNIIILNAGNKDEKMEKIOEKLEQILPTSGNKETRGPNOSVNDROSL 1170
QY 153 WTEHSHEIWEMLCAL-----VKIGAKKDDFTEN-----YGYNNVK 188
Db 1171 WDRIAEHYHGWNCALTYKDDNGILKGYVKKPKQIENPEKLMNETTKPKDEKYOYQAK 1230
QY 189 FSDKS-----TLEEFARPOFLRLTEWYDDYCYTROKYLRDVOEKRSN--- 234
Db 1231 LEDESGERPDSASAGTKLTDIKRPYFRYLEWGENCKKRTMLKIKEDCYKNGR 1290
QY 235 ---DQKLC---DTE-----CNKKCEDYKYKKKKKE-----WIPDKYYK 268
Db 1291 CSGGLKCKNEIVIDKEKIFGLCLCPTCARHCFYKRWINTKRDENKOSNAYSEOKKYE 1350
QY 269 DERD 272
Db 1351 EEND 1354

RESULT 15

T18396
erythrocyte membrane protein variant 2 - malaria parasite (Plasmodium falciparum) (fragm
C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18396
R:Baruch, D.I.; Pasloske, B.L.; Singh, H.B.; Bi, X.; Ma, X.C.; Feldman, M.; Taraschi,
Cell 82, 77-87, 1995
A:Title: Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen a
A:Reference number: 218925; MUID:95330812; PMID:7541722
A:Accession: T18396
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1729 <BAR>
A:Cross-references: EMBL:U27339; NID:9914920; PID:9914921; PIDN:AAA89134.1
C:Genetics:
A:Gene: EMP1
A:Note: var-2

Query Match 19.4%; Score 299.5; DB 2; Length 1729;
Best Local Similarity 22.6%; Pred. No. 1.6e-14;
Matches 99; Conservative 47; Mismatches 121; Indels 171; Gaps 15;

QY 2 CHPKNSNGYPMOCGNINLVEDPR-----VCMPPRROKICVHFL----- 41
Db 854 CLKXGKNAPISMKC-----IPSDTKSVATGSDTGSISICVPRRRRLVYGGKLHDMAGG 909
QY 42 -----ANDN-----BIKKLSQVNLKEAFKSAAEFPFS 71
Db 910 ETTEAKSOETSOGCKTPSGNESSPESEKLPQGPPTETPKETEPSSILHAFVSPRLRFLP 969
QY 72 WYYSK-----DGEENELDKELKEGKIPPAFLRSFYFGDYDF 112
Db 970 WAKKEQMKAOHAGCATGLQLPGYTVDDSDPPQTQLRGNLINDFLKMFYTLGDYRDI 1029
QY 113 LFGTD-----ISKHGBGSKLKEQIDSLFKNQDOKSPNGK----- 147
Db 1030 CIGGDRDIVGDTIVSITGSEYKKIKSILIEGFLKKQYVTSPPRDTSSRPVHPQTSVE 1089
QY 148 --TROEMTSHETWEMLCALV-----KLT-----GAK--- 174
Db 1090 KTPQQTWEANGPHIWNMICALTYEDSAGIOPPOKVEDADKYLEKLPNTANGIKMYL 1149
QY 175 KDDFENTGYNNVFKSDKST-----TLEEFARPOFLRLTEWYDDYCYTROKYLRDQV 228
Db 1150 KEDNTSSAMPTSSSSSSSSGSDNPINTPKLIEVEIPTFRYRLHEWQNCCKERMKRLKOIY 1209
QY 229 EKCKSND-----QLKC-----DT-----EKNKCEDYKYKMK 255
Db 1210 KCKVGENGYGRKQKTPQSCCYCEDCEDQSKSYDTVADLECPKCAKHCRWYKKMIE 1269
QY 256 KKK-EMIPQDKYKDERD 272
Db 1270 KKDEFTDEKAFPRKOD 1287

Search completed: April 28, 2003, 10:32:56
Job time : 16.4383 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:46 ; Search time 2.65193 Seconds

(Without alignments)
4316.649 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554

Perfect score: 1543
Sequence: 1 DCHPKNSNGTPDMCCGNIN.....KKWIPDCKYKDEKRRF 276

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	274	17.8	1435	1	EBAL_PLAFC
2	236	15.3	1153	1	PVDB_PLAFC
3	223	14.5	1070	1	PVDB_PLAFC
4	218.5	14.2	1073	1	PVDB_PLAFC
5	214.5	13.9	1070	1	PVDB_PLAFC
6	105.5	6.8	346	1	YE77_METUA
7	103.5	6.7	677	1	RPOC_SPIOL
8	99	6.4	3660	1	DMD_CHICK
9	98.5	6.4	1341	1	RPA1_METUA
10	97.5	6.3	680	1	RPOC_TOBAC
11	95.5	6.2	678	1	GARP_PLAFC
12	95	6.2	4427	1	PKSL_BACSU
13	94.5	6.1	446	1	TB11_NOTOC
14	94.5	6.1	1178	1	MNNA_YEAST
15	94	6.1	1081	1	SYT_TERTH
16	92.5	6.0	441	1	Y107_METUA
17	92.5	6.0	450	1	TBB2_GAEL
18	92	6.0	451	1	TBB2_GAEL
19	91	5.9	434	1	TBBQ_HUMAN
20	91	5.9	444	1	TBB1_HUMAN
21	91	5.9	444	1	TBB5_HUMAN
22	91	5.9	444	1	TBB7_CHICK
23	91	5.9	444	1	TBBX_HUMAN
24	91	5.9	450	1	TBB_BOMO
25	91	5.9	617	1	Y237_BUCAL
26	91	5.9	1230	1	TBBE_YEAST
27	90.5	5.9	445	1	TBB_PSEAM
28	90.5	5.9	446	1	TBB6_CHICK
29	90	5.8	467	1	TBB1_PHYPO
30	90	5.8	1391	1	MST2_DROHY
31	90	5.8	2351	1	FAB_HUMAN
32	89.5	5.8	680	1	RPOC_ARATH
33	89.5	5.8	1164	1	BAC_STRAG

34	89	5.8	444	1	TBB_ONCGI	P41387 onchoerca
35	89	5.8	446	1	TBB_GALA	P05304 giardia lam
36	89	5.8	712	1	ARS2_GAEL	O96615 caenorhabdi
37	88.5	5.7	441	1	TBB7_GAEL	O17299 caenorhabdi
38	88.5	5.7	441	1	TBB7_GAEL	P12456 caenorhabdi
39	88.5	5.7	449	1	TBB4_CHICK	P09652 gallus gall
40	88.5	5.7	450	1	TBB3_MOUSE	O96d7 mus musculu
41	88.5	5.7	450	1	TBB4_HUMAN	Q13509 homo sapien
42	88.5	5.7	734	1	YKR2_GAEL	P34308 caenorhabdi
43	88.5	5.7	944	1	NUR1_YEAST	P32380 saccharomyc
44	88	5.7	442	1	TBB_STYLE	P1857 stilyonchia
45	88	5.7	448	1	TBB1_BRUPA	P18241 brugia paha

ALIGNMENTS

RESULT 1	ID	EBAL_PLAFC	STANDARD:	PRT: 1435 AA.
AC	P19214	EBAL_PLAFC		
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	01-FEB-1996 (Rel. 33, Last annotation update)			
DE	Erythrocyte-binding antigen EBA-1/5.			
OS	Plasmodium falciparum (Isolate Camp / Malaysia).			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.			
OX	NCBI_TaxID=5835;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90377299; PubMed=2204835;			
RA	Slm B.K.L.;			
RT	*Sequence conservation of a functional domain of erythrocyte binding			
RT	antigen 175 in Plasmodium falciparum.*;			
RL	Mol. Biochem. Parasitol. 41:293-296(1990).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement. (See http://www.isb.ch/announce/			
CC	or send an email to license@isb.slb.ch .)			
DR	EMBL; X52524; CAA36756.1; -			
DR	PIR; S11561; S11561.			
KW	Antigen.			
FT	DOMAIN 159 1104			
FT	ESSENTIAL FOR BINDING TO			
FT	ERYTHROCYTES.			
FT	E -> V (IN STRAINS FCR-3 AND ITG).			
SQ	VARIANT 1031 1031			
SQ	SEQUENCE 1435 AA; 167389 MW; 3244309021BIC3D6 CRC64;			
Query Match	17.8%; Score 274; DB 1; Length 1435;			
Best Local Similarity	29.2%; Pred. No. 1.3e-13;			
Matches	76; Conservative 43; Mismatches 93; Indels 48; Gaps 11;			
QY	14 MOCNTLVDPVPCMPRRQKLCVHPLANDNEIKLQSOVNLKEAFIKSAAEFFSWY 73			
DB	474 WECKNPIYISTKVCVPRQELC--LGNIDRIYD-KNLMKEHLLAIAIESRLIKR 529			
QY	74 YYSKSGEGNELKEKGIPIPAFLRSMFYTGDFGFDISGHEGSKLQID 133			
DB	530 KYRNKD-----DREV-----CKIINKTADIDIDIGGTDMNDL-SNRKLVGKIN 573			
QY	134 SLKNGDKSPNKT-ROEWTEHSHIWMALCALVIAKKDDPLENGYNNVRSOK 192			
DB	574 TNSKIYHRNKNKLFEDEMWKYIKKDVNVT-----SWFKDK 612			
QY	193 STLLIEFAKR-POFLRWLTETDYCYTRQKRYLKDVOEKKSNDOELKC-DTECNKCEDY 250			
DB	613 TVCKEDDIENIPQFRRFSSMGDDYCDQDKTKMETLKVCK-----EKRCEDNCKSKNSY 669			
QY	251 VKTM-KKKKEMIPDCKYKD 269			

QY	2	CHPKNSNGYVPMOCGINLVEDPPVCAHPRRQKLCVHFLAN-----DNEIKLIQS	52
DB	214	CNDKR-KRGERDMD-PAEKDICISVRYOZLCKMGLTGLVNNTRTHSHNDTFL--	265
QY	53	OVNLAEVRIKSAARATFFSWYYKSKDOEGNDLEDEKGRIPPAFLSMFPTEDYDF	112
DB	266	KLNLRKRLMYDAV-----EGDILLRK--NNYQTNKEFCQDIRGCDPEDI	310
QY	113	LEGTDISKHGEGSKLKREQDLSLFKNQDOKSPNGTRQREWTEHSHETMEALCALVIG	172
DB	311	IMGTM-EGIGYSQVVENLRQVE-GTDEKAK--QDRKQNNSEKHIIMRAMFESI----	362
QY	173	AKKDPFTENYGVNNVKSFDKSTLLEFPAKRQFLWLTWMDYCYTTRQKYLAKDQECCK	232
DB	363	--RSRLKRF-----TWICKKDVTLK--VEPDIYMIREWGRDYMYSKLPRKQGLNEKCA	413
QY	233	S---NDOLKCDTE--CNKKCEDYVVM--KKKKEM	260
DB	414	SKLYYNNMAICMLPLCHDACKSYDDMIIRKKQW	447
RESULT 5			
PVDG--PLAKN		STANDARD:	PRT: 1070 AA.
ID	PVDG_PLAKN		
AC	P50494:		
DT	01-OCT-1996	(Rel. 34, Created)	
DT	01-OCT-1996	(Rel. 34, Last sequence update)	
DT	01-OCT-1996	(Rel. 34, Last annotation update)	
DE	Duffy receptor, gamma form precursor (Erythrocyte binding protein).		
OS	Plasmodium knowlesi.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5850;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=92357776; PubMed=1496004;		
RA	Adams J.H., Sim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.;		
RT	"A family of erythrocyte binding proteins of malaria parasites."		
RL	Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).		
CC	-1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP DETERMINANT.		
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-1- SIMILARITY: HIGH, TO P.VIVAX DUFFY RECEPTOR.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use, by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/or-send-an-email-to-license@isb-sib.ch).		
DR	EMBL: M90695; AAA29604.1;		
KW	Malaria; Receptor; Glycoprotein; Signal; Transmembrane; Multigene family.		
KW	SIGNAL	1	21
FT	CHAIN	22	1070
FT	DOMAIN	22	1003
FT	TRANSMEM	1004	1025
FT	DOMAIN	1026	1070
FT	CARBOHYD	134	134
FT	CARBOHYD	179	179
FT	CARBOHYD	676	676
FT	CARBOHYD	743	743
FT	CARBOHYD	785	785
FT	CARBOHYD	936	936
FT	SITE	279	281
SO	SEQUENCE	1070 AA;	120931 MW; 70356881bcl1b50 CRC64;
Query Match	13.9%;	Score 214.5;	DB 1; Length 1070;
Best Local Similarity	23.9%;	Pred. NO. 3.9e-09;	

Matches 68; Conservative 48; Mismatches 91; Indels 77; Gaps 14;

QY 2 CHPKNSNGYPMQCGNINIVEDPRVCPMPRROKLCVHFLAN-----DNEIKLQ 51
 Db 214 CNDRK-KRGEDMDCPT-----ENDVCIPDRKRIOLCMEITNLVDYDTHFHSIDIFKRSY 267
 QY 52 SQVUL-----KRAFKSAAEETFEFSWYKSKDGEINDELKCKGKIPPAFLRSMFY 104
 Db 268 SRRLIYDVGGRRDLTLK-----KYNMYSEDLCKDK-----W 301
 QY 105 TFGDRDFLEGTDISKGGHSGSKLKEQIDSLFKNGQOKSNGKTRQEWTEHSHEITWAM 164
 Db 302 SLDPFGDIIINGTDM-BEIGSLVYVANNKSIPTGTSAELD---RKMWNDRKDKDIWKAM 357
 QY 165 LCLAVKIGAKKDDFTENYGVNNYKFSDDKSTTLEFAKRPQFLMTLEWYDYCYTROXYL 224
 Db 358 ILSVKE-----KNYSAMNCK-EDVQIKVE-----PQIYMIEMGMDYASEREQR 403
 QY 225 KDVOEKCKSNOLKCDP-----ECNKKCEYVYKMK-KKKEM 260
 Db 404 RKLNEKE--DKLYSTMLICTLPPCNMACKSYDEWITGKKOW 445

RESULT 6
YE77_METJA

AC 058872; STANDARD; PRT; 346 AA.

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ1477.
 GN MJ1477.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 CC Methanocaldococcaceae; Methanocaldococcus.
 OK NCBI_TaxID=2190;
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=9637999; PubMed=6868087;

RA Bult C.J., White O., Olsen G.V., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kervaege A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.L.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Uitterlinden T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Blank H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073(1996).

CC -1- SIMILARITY: TO D.RADIOURANS DR0705 AND T.MARITIMA TM1410.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U67588; AAB99489.1; -

DR TIGR: MJ1477; -

DR InterPro: IPR004352; TM1410-related.

DR Pfam: PF03537; DUF297.1.

DR PRINTS: PR01545; THEMAYEL0DUF.

KM Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 16 36 POTENTIAL.

SC SQUENCE 346 AA; 40620 MW; A71B3D69A956BCE CRC64;

Query Match 6.8%; Score 105.5; DB 1; Length 346;
 Best local Similarity 22.1%; Pred. No. 0.31; 93; Indels 97; Gaps 14;

QY 43 NDNEIKLQSQVNLKRAFKSAAEETFEFSWYKSKDGEIN-----ELDKELRKGKIP 96
 Db 63 NSNNLKFMA-YOLOADIDEIANSFTLLIYIDYSDGDNENKRYSEELKTKKAKKIP 121
 QY 97 AFLRSFETFGDYRDLFTGTD-----ISKHGECSK-LKBOI 132
 Db 122 AIT-----SIGEAEDRFVNDNEMLKNPPKWLGDENPEWEGCYAVKVMPEWKILFSYL 176
 QY 133 DSLFKNGDKSPNGKTRQ-EMWTEHSHEITWAMLCALYKIGAKKDDT-----E 180
 Db 177 DKTIQGGFCGYVLDVYDEEYEAENGID-----EDFAKEMIKFIYBIS 220
 QY 181 NGYNNVKS-----DKSTTLEFAKRPQFLMTLEW-YDYCY--TROXYLKVOEKCK 232
 Db 221 NCRNKTNSFTIIPONGERLLEYDKHGLMTVSGMAVEDLYDGVGQKTEEELNERIK 280
 QY 233 SNDQKCD-----TECN--KKGCD-----YKYMCK 256
 Db 281 LLDKVKDGGKFLVYVDYDDGKTINENLKRVEDFINKSLDKGYVYVAK 329

RESULT 7
RPOC_SPIO

AC P11705; STANDARD; PRT; 677 AA.

DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6).
 GN RPOC1.
 OS Spinacia oleracea (Spinach).
 OC Chloroplast.
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC Caryophyllales; Caryophyllales; Chenopodiaceae; Spinacia.
 OK NCBI_TaxID=3562;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=8831693; PubMed=3045324;
 RA Hindson G.S., Holton T.A., Whitfield P.R., Bottomley W.;
 RT "Spinach chloroplast rpoB genes encode three subunits of the
 RT chloroplast RNA polymerase.";
 RL J. Mol. Biol. 200:639-654(1988).
 RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-cv. Giant d'hyver, and cv. Monato;

RX MEDLINE=21187424; PubMed=1192076;

RA Schmitz-Limweder C., Maier R.M., Alcaraz J.-P., Cottet A.,
 RA Hermann R.G., Mache R.;
 RT "The plastid chromosome of spinach (Spinacia oleracea): complete
 RT nucleotide sequence and gene organization.";
 RL Plant Mol. Biol. 45:307-315(2001)

CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA)(n).

CC -1- SUBUNIT: IN CHLOROPLAST THE RNA POLYMERASE IS COMPOSED OF FOUR
 CC SUBUNITS: ALPHA, BETA, BETA', AND BETA".

CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AJ400848; CAB88716.1; -

DR PIR: B29959; B29959.

DR HSSP: Q9KRW6; 1HOM.
DR InterPro: IPR000722; RNA_pol_A.
DR Pfam: PF00623; RNA_pol_A.1.
DR Transference: Transcription; DNA-directed RNA polymerase; Chloroplast.
SQ SEQUENCE 677 AA; 78149 MW; D618E17A829CEBD CRC64;

Query Match
Best Local Similarity 24.6%; Score 103.5; DB 1; Length 677;
Matches 51; Conservative 22; Mismatches 53; Indels 81; Gaps 13;

QY 16 CGNINLV-----EDPRVCMPPRRQKLCVHFLANNEIKLOSYNLKAFFKSAAEFFS 71
DB 71 CGNRYVIGDEKEDPKRC-----ECCGVEFY--DSRIKRYGM-----GYIKLACPVTHV- 116
QY 72 WYVYKSKDGEENDELK--EGKI-----PPAFLR-----S 101
DB 117 WYKRLRPSYIANFLDKPLKELEGLVCDFFSAPRIAKKPFRLRGFLFEYEQSKYSIP 176
QY 102 MFTY---FGDYRDFLPGTQDSKSGHSGSKKEQ-----IDSLFKNDQKSPNCKTR 149
177 LFTTGGFDTFRN---REISTGAG--AIREQLADLDLRTIIDYSFAEWEKELGEGSTG 229
150 QEWTEHSEHWMLCALVKIGAKKD 176
DB 230 NEMEDR-----KVGRRKD 242

RESULT 8
DMD_CHICK
ID DMD_CHICK STANDARD; PRT; 3660 AA.
AC P11533;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dystrophin.
GN DMD.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8908331; PubMed=3062582;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "Nucleotide sequence of chicken dystrophin cDNA";
RL Nucleic Acids Res. 16:11815-11815(1988).
RN [2]
SEQUENCE FROM N.A.
TISSUE=Muscle;
MEDLINE=89210800; PubMed=3072195;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "The chicken dystrophin cDNA: striking conservation of the C-terminal
coding and 3' untranslated regions between man and chicken.";
RL EMBO J. 7:4157-4162(1988).
CC -1- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE
PLASMA MEMBRANE.
CC -1- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FILABIN,
ABP-120, ABP-180, OR BETA-FODRIN).
CC -1- SIMILARITY: CONTAINS 2 CALPONTIN-HOMOLOGY (CH) DOMAINS.
CC -1- SIMILARITY: CONTAINS 22 SPECTRIN REPEATS.
CC -1- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: X13369; CAA31746.1; -.
DR PIR: S02041; S02041.
DR HSSP: P46939; 1BHD.
DR InterPro: IPR001589; Actbind actin.
DR InterPro: IPR001725; Calpoin-like.
DR InterPro: IPR002017; Spectrin.
DR InterPro: IPR002349; WW.
DR InterPro: IPR001202; WW_Rsp5_WWP.
DR InterPro: IPR000433; Znf_ZZ.
DR Pfam: PF00307; CH; 2.
DR Pfam: PF00435; spectrin; 22.
DR Pfam: PF00569; Z2; 1.
DR PRINTS: PR00403; WMDOMAIN.
DR SMART: SM00033; CH; 2.
DR SMART: SM00456; SPEC; 21.
DR SMART: SM00291; ZNF_ZZ; 1.
DR PROSITE: PS00019; ACTININ_1; 1.
DR PROSITE: PS00020; ACTININ_2; 1.
DR PROSITE: PS50021; CH; 2.
DR PROSITE: PS01159; WW_DOMAIN_1; 1.
DR PROSITE: PS50020; WW_DOMAIN_2; 1.
DR PROSITE: PS01357; ZF_ZZ_1; 1.
DR PROSITE: PS50135; ZF_ZZ_2; 1.
DR Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
KW Repeat; Zinc-finger.
FT DOMAIN 1 244 ACTIN-BINDING.
FT DOMAIN 19 123 CH 1.
FT DOMAIN 138 241 CH 2.
FT REPEAT 341 449 SPECTRIN 1.
FT REPEAT 450 558 SPECTRIN 2.
FT REPEAT 561 669 SPECTRIN 3.
FT REPEAT 721 830 SPECTRIN 4.
FT REPEAT 832 936 SPECTRIN 5.
FT REPEAT 945 1047 SPECTRIN 6.
FT REPEAT 1050 1156 SPECTRIN 7.
FT REPEAT 1159 1265 SPECTRIN 8.
FT REPEAT 1268 1369 SPECTRIN 9.
FT REPEAT 1470 1570 SPECTRIN 10.
FT REPEAT 1573 1678 SPECTRIN 11.
FT REPEAT 1681 1782 SPECTRIN 12.
FT REPEAT 1879 1981 SPECTRIN 13.
FT REPEAT 2013 2103 SPECTRIN 14.
FT REPEAT 2106 2211 SPECTRIN 15.
FT REPEAT 2214 2321 SPECTRIN 16.
FT REPEAT 2472 2574 SPECTRIN 17.
FT REPEAT 2577 2683 SPECTRIN 18.
FT REPEAT 2686 2799 SPECTRIN 19.
FT REPEAT 2802 2904 SPECTRIN 20.
FT REPEAT 2906 2928 SPECTRIN 21.
FT REPEAT 2931 3037 SPECTRIN 22.
FT REPEAT 3052 3085 WW.
FT ZN_FING 3304 3351 Z2-TYPE.
FT VARIANT 1171 1171 MISSING.
FT VARIANT 1171 1171 MISSING.
FT VARIANT 1869 1869 Q -> H.
FT VARIANT 1885 1885 K -> R.
SQ SEQUENCE 3660 AA; 422874 MW; 85493DAF6D5B6DA4 CRC64;

Query Match
Best Local Similarity 18.7%; Score 99; DB 1; Length 3660;
Matches 53; Conservative 49; Mismatches 90; Indels 92; Gaps 11;

QY 45 NEIKLOSQVNLKFAFFKSAAEFFFSMYVYKSDGSGNEL-----DKELKGIIPAF 99
DB 1861 NALKDLASQRRK-----ALEISHQVYQYR---QADDLMTWLDIETKIASLP----- 1906
QY 100 RSMFYTGDRDFLFGDLSKSGHSGSKLKEQDLSLFKNGDQKSPNCKTRQEWTEHSE 159
DB 1907 -----DHKD---EQKLKEIGELKKEDLVANVRQAEERLSKDGAAK----- 1945
QY 160 IWEAMLCALVKIGAKKDPTFENYGVYNNVAFSDKSTLLE----- 198

Db 1946 ---AVEPVLVQSKRWDESKFAQFRRLNVAIOVLDTFTVMTESVETVTPSTY 2002
 QY 199 FAKRPFLEMLTWDYCYTRQKYLKDVQCKNDLQKOTECNKCCEYVAKMKKK 258
 Db 2003 LAEIIQLQALSE-VEERLNSPVLAQKCEDLLKQOEELKNIKIDGLGRIGLIDHHSK 2061
 QY 259 -----EWIPDKYKXDER---DKR 275
 Db 2062 TPALASAPRETANTQDKLTOLNSQWEKYNKMYRDRQARFDRKSK 2105
 RESULT 9
 ID REPAL_METUA STANDARD; PRT; 1341 AA.
 AC 058445;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA-directed RNA polymerase subunit A' (EC 2.7.7.6) [Contains: Mja
 RPOA1 Intein (Mja RpoA' Intein)].
 RPOA1 OR MJ1042.
 Methanococcus jannaschii.
 Archaea; Euryarchaeota; Methanococci; Methanococcales;
 Methanocaldococcaceae; Methanocaldococcus.
 NCBI_TaxID=2190;
 [1]
 SEQUENCE FROM N.A.
 STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bilt C.V., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 Sulten G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 Kervatag A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Gladen A.,
 Scott J.L., Geoghegan N.S.M., Meldrum J.F., Fuhmann J.L., Nguyen D.,
 Uffebach T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 Science 273:1058-1073(1996).
 RL Science 273:1058-1073(1996).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 {RNA}(N).
 CC -1- COFACTOR: ZINC.
 CC -1- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
 A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
 FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
 CC -1- SIMILARITY: THE COMBINED A+A' SUBUNITS CORRESPOND TO THE A
 SUBUNITS OF EUKARYOTIC RNA POLYMERASES I, II AND III AND TO THE
 EUKARYOTIC BETA' SUBUNIT.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 or send an email to license@1sb-sib.ch).
 CC DR EMBL; U67547; AAB99046.1;
 DR TIGR; MJ1042;
 DR InterPro; IPR003586; Hedgehog_hntc.
 DR InterPro; IPR003587; Hedgehog_hntn.
 DR InterPro; IPR002203; Intein.
 DR InterPro; IPR004042; Intein_endonc.
 DR Pfam; PF00623; RNA_pol_A_1.
 DR SMART; SM00305; Hntc; 1.
 DR SMART; SM00306; Hntn; 1.
 DR PROSITE; PS50818; INTEIN_C_TER; 1.
 DR PROSITE; PS50819; INTEIN_ENDONUCLEASE; 1.

DR PROSITE; PS50817; INTEIN_N_TER; 1.
 KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
 KW Autocatalytic cleavage; Protein splicing; Complete proteome;
 KW Zinc-finger.
 FT CHAIN 1 460
 FT CHAIN 461 911
 FT CHAIN 912 1341
 FT ZN_FING 62 105
 SQ SEQUENCE 1341 AA; 152780 MW; 4D98546ED552A12 CRC64;
 Query Match 6.4%; Score 98.5; DB 1; Length 1341;
 Best local Similarity 19.8%; Pred. No. 5.1;
 Matches 71; Conservative 51; Mismatches 111; Indels 125; Gaps 16;
 QY 3 HPRKNSGYPDWQCGNINIVEDPRVCMPPRRQKLCVHPLANDNEKRLQSOVNLKEAFIK 62
 Db 533 HPRYTNGRR--RCGELK-VGDEVIIIP-----NDFPFEDNRYVDEEKIK 577
 QY 63 SAALTFESWYTKSKDGNELDEKELKGIIPAFRLSMYTFEGDPLFGTDISKGH 122
 Db 578 KVINNGT--YK-----NKINELDKRLIP-----LYNDQKASILARIYGHVM 621
 QY 123 GEGS-----KLKQIDSLFNGDQ-----KSPNCKTQEW 152
 Db 622 GDGSLINNNKNSRVYFRGDIEDLTKIKEDLKELDYDGEIKELHGEETIDYNGKKRIIK 681
 QY 153 WTEHSEIWEALCALVR-IGAKKDDFTN-YGVNN----- 186
 Db 682 GKGYFEFRKKSICILALACVGGDKTKMYGIPNMTKTPKIKKEFLSAYGSELT 741
 QY 187 VFSKSTLLEFARPPQELMLTWDYCYTRQKYLKDVQCKNS-NDQLKDECNK 245
 Db 742 PRIRHSGSEFELSK---IAKIEIFDE---DREIKIDKELKFGELKYRVEGN 793
 QY 246 -----KEDYKYTKKKKKEN-----IPDKYKDERDK 274
 Db 794 LRKQYKTKYVVASIYNHKEFFGRIGTYANKRETLARYAVETLLMEKLIKRNRIK 851
 RESULT 10
 ID RPOC_TOBAC STANDARD; PRT; 680 AA.
 AC P12116;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6).
 GN RPOC1.
 OS Nicotiana tabacum (Common tobacco).
 CC Chloroplast.
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 CC NCBI_TaxID=4097;
 CC [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-cv. Bright yellow 4;
 RA Shinozaki K., Ohme K., Tanaka M., Wakasugi T., Hayashida N.,
 Matsubayashi T., Zaita N., Chunwongse J., Obozaki J.,
 Yamaguchi-Shinozaki K., Ohno C., Torazawa K., Meng B.-Y., Sugita M.,
 Deno H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A.,
 Tohoh N., Shimada H., Sugitara M.;
 RT "The complete nucleotide sequence of the tobacco chloroplast genome:
 its gene organization and expression.";
 RL EMBO J. 5:2043-2049(1986).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 {RNA}(N).
 CC -1- SUBUNIT: IN CHLOROPLAST THE RNA POLYMERASE IS COMPOSED OF FOUR
 SUBUNITS: ALPHA, BETA, BETA', AND BETA".
 CC -1- SUBCELLULAR LOCATION: Chloroplast.

CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z00044; CAA77411.1; ALT_INIT.
 CC PIR: A05033; A05032.
 CC PIR: A05033; A05033.
 CC HSSP: Q9K006; 1HQW.
 CC InterPro: IPR000722; RNA_pol_A.
 CC Pfam: PF00623; RNA_pol_A; 1.
 CC Transferrase; Transcription; DNA-directed RNA polymerase; Chloroplast.
 CC SEQUENCE 680 AA; 78427 MW; 6D3FE9BC8D3A291 CRC64;
 SO
 Query Match 6.3%; Score 97.5; DB 1; Length 680;
 Best Local Similarity 23.7%; Pred. No. 2.8;
 Matches 49; Conservative 25; Mismatches 52; Indels 81; Gaps 13;
 QY 16 CGNINLV-----EDPRYCMPPRROKLCVHFLANDNEIKKQSOVNAKEAFIKSAAEFFS 71
 DB 70 CGNRYVIGDEKEDPKFC-----EQGVEYV--DSRIKRYOM-----GYIKLACPVTHV- 115
 QY 72 WYVYKSKDEGNEIDKELK--EKGK-----PPAFLR-----S 101
 DB 116 WYKRLRPSTYIANLDKRLKLEGLVYCDPSFARPIKPKFLRLGLEFEIOSWKYSIP 175
 QY 102 MEFTY---FGDYRDLFTGTDISKHSGSKLKEO-----IDSLFKNDQKSPNGKTR 149
 DB 176 LFFETGQFDFFRN-----REISTGAG--AIRQLADLDLIILNLSIVMEWEIELEGHTG 228
 QY 150 QEWTEHSHETWEAMCALVKGAKKD 176
 DB 229 NEMEDR-----KVRKRD 241
 RESULT 11
 GARP PLAAF STANDARD; PRT; 678 AA.
 ID GARP PLAAF
 AC P13816;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Glutamic acid-rich protein precursor.
 GARP.
 Plasmodium falciparum (isolate FC27 / Papua New Guinea).
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=5837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89040048; PubMed=2903445;
 RA Trigila T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,
 RA Kemp D.J.;
 RT "Structure of a Plasmodium falciparum gene that encodes a glutamic
 RT acid-rich protein (GARP).";
 RL Mol. Biochem. Parasitol. 31:199-202(1988).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J03998; AAA29605.1;
 CC PIR: A54514; A54514.
 CC Repeat; Malaria; Antigen; Signal.
 FT SIGNAL 1 25

FT CHAIN 26 678
 FT DOMAIN 120 164
 FT DOMAIN 372 416
 FT DOMAIN 417 441
 FT DOMAIN 576 604
 FT DOMAIN 605 653
 FT DOMAIN 654 663
 SO SEQUENCE 678 AA; 80551 MW; 2ABF85606496EA9E CRC64;
 SO
 Query Match 6.2%; Score 95.5; DB 1; Length 678;
 Best Local Similarity 20.8%; Pred. No. 4;
 Matches 55; Conservative 42; Mismatches 92; Indels 75; Gaps 10;
 QY 43 NDNEIKKLOSQVNLKEAFIKSAAEFFSWYVYKSKDGEENIDKELKEGKIPAPFLRSK 102
 DB 113 NENSVKRRKKDKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKR 154
 QY 103 FYTFGDRDLFTGTDISKHSGSKLKE--QIDSLFRNGQKSPNGKTRQEWMTSHET 160
 DB 155 -----EKKHKDKRKKRKNSEVSLYRTGQHKPRNA-----TEHGEEN 191
 QY 161 WEAMCALVKGAKKD-----DETFENVG-----YNNVFRSDKSTYLE--EFA 200
 DB 192 LDEAMVSEINNAOGGLLSSPYQREOGCGIISVHETSNDTKDNKINISEDKEDH 251
 QY 201 KRQFLRMT-----TEWYDYCYTRQRYKLDVQEKCKSNDQKCDTECKNKCEDYV 251
 DB 252 QQEMMLKTLDDKRRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKR 309
 QY 252 KYKKKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKR 275
 DB 310 KKKKKKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKR 332
 RESULT 12
 PKSL_BACSV STANDARD; PRT; 4427 AA.
 ID PKSL_BACSV
 AC Q03470;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative polyketide synthase pksl (PKS).
 GN PKSL OR PKSX OR PKSA OR OUTG.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
 NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / P81424;
 RX MEDLINE=93345824; PubMed=8344529;
 RA Scotti C., Piatelli M., Guzzoni A., Perani P., Tognoni A., Grandi G.,
 RA Galizzi A., Albertini A.M.;
 RT "A Bacillus subtilis large ORF coding for a polypeptide highly
 RT similar to polyketide synthases.";
 RL gene 130:65-71(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunze F., Ogasawara N., Moszer I., Albertini A.M., Allion G.,
 RA Azevedo V., Bertero M.G., Bessières P., Bolotin A., Borchert S.,
 RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
 RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Ertlan K.D., Errington J., Fabret C., Ferrari E., Rouger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Gilm S.T., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Gutseppl G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaere-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter T., Kraschinsky G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Moone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetle D., Portwoll S., Prescott A.M.,
RA Prescan E., Puc P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Schenck S., Schreier R., Scottone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shih B.S., Soldo B.,
RA Sotokini A., Taconi E., Takagi T., Takahashi H., Takemori K.,
RA Takeda V., Uchiyama S., Vandenbol M., Vanlier F., Vassarotti A.,
RA Tosi A., Wambuit R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*,"
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE OF 3619-4427 FROM N.A.
RR STRAIN=168 / PBL424;

Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.

-1- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
SECONDARY METABOLISM

-1- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTHETINES
(POTENTIAL).

-1- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: 214098; CAA78479.1; -
DR EMBL: 011039; AAA85145.1; -
DR EMBL: 299113; CAA13602.1; -
DR EMBL: 235133; CAA84504.1; -
DR PIR: S25021; S25021.
DR Subtilist; B610698; pKSL.
DR InterPro: IPR002198; ADH_short.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR003880; ppantte_attach.
DR Pfam: PF00106; ash_short; 1.
DR Pfam: PF00109; ketoacyl-synt; 4.
DR Pfam: PF00550; pp-binding; 5.
DR Pfam: PF02801; ketoacyl-synt; 4.
DR PROSITE: PS00012; PHOSPHOPANTHETINE; 5.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE: PS0075; ACP_DOMAIN; 5.
DR Transferrase; Acyltransferase; Antibiotic biosynthesis; NADP;
DR Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
DR Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT DOMAIN 211 280
FT DOMAIN 382 759
FT DOMAIN 937 1115
FT DOMAIN 1409 1602
FT DOMAIN 1687 1759
FT DOMAIN 1876 2253
FT DOMAIN 2491 2560
FT DOMAIN 2632 2701
FT DOMAIN 2823 3182
FT DOMAIN 3575 3776
FT DOMAIN 3854 3923
FT DOMAIN 4019 4373
FT BINDING 243 243
FT BINDING 1723 1723
FT BINDING 2523 2523
FT BINDING 2664 2664
FT BINDING 3886 3886
FT BINDING 4427 4427
FT BINDING 493398 MW; 961251E561AB9F2 CRC64;
SQ SEQUENCE

Query Match 6.2%; Score 95; DB 1; Length 4427;
Best Local Similarity 24.0%; Pred. No. 37;
Matches 43; Conservative 41; Mismatches 69; Indels 26; Gaps 8;

OY 110 RDLFTGDSKSGGSESKLEQDLSFKNGDDKSPGKTRQMTWHEHSHETAEALCALV 169
DB 1648 RALHET-----GKGKQSNIKKSSLNQAEKADMGIDEITLNEKAKENFYKQVLSVT 1703
OY 170 KIGAKKDF--TENGYNVNFESDKSTLEE---FAK---RPOFLMLTWTYDYCY 218
DB 1704 KIPAGGIDAEADLEVDYGDISIMIMHTGQLEKVFSLSTLEFFEDYRSLRYRID--- 1760
OY 219 TRQYKVDYQ--EKKSNDQKCDPECKKCDYKAKKKKEWIP--ODKYKDERDK 274
DB 1761 SRREKLIDLTGFEETGKPSVERSEPE---KQELIPVIRKSGFLPDCKEQVREKE 1814

RESULT 13
TB1 NOTCO
ID TB1 NOTCO STANDARD: PRT; 446 AA.
AC P36221
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Tubulin beta-1 chain.
GN TUBB1.
OS Notothenia coriiceps neglecta (Black rockcod) (Yellowbelly rockcod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Teleostei; Neoteleostei;
OC Acanthomorphi; Acanthopterygii; Perciformes; Perciformes;
OC Nototheniidae; Nototheniidae; Notothenia.
OX NCBI_Taxid=8209;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-40.
RC TISSUE=Brain;
RX MEDLINE=93223253; PubMed=8467523;
RA Detric H.W. III, Parker S.K.;
RT "Divergent neural beta tubulin from the Antarctic fish *Notothenia*
RT *coriiceps neglecta*: potential sequence contributions to cold
RT adaptation of microtubule assembly";
RL Cell Motil. Cytoskeleton 24:156-166(1993).
CC -1- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT
CC BINDS TWO MOLES OF GTP. ONE AT AN EXCHANGEABLE SITE ON THE BETA
CC CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.
CC -1- SUBUNIT: DIMER OF ALPHA AND BETA CHAINS.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: L08013; AAA49393.1; -
DR EMBL: S57698; AAB26110.1; -
DR PIR: A48407; A48407.
DR InterPro: IPR002453; Beta_tubulin.
DR InterPro: IPR000217; Tubulin.
DR InterPro: IPR003008; Tubulin_ftsz.
DR Pfam: PF00091; tubulin; 1.
DR PRINTS: PF01161; TUBULIN.
DR PROSITE: PS00227; TUBULIN_1;
DR PROSITE: PS00228; TUBULIN_B_AUTOREG; 1.
DR Microtubules; GTP-binding.
FT NP_BIND 140 146
FT NP_BIND 446 AA; 49806 MW; B8CE37298707D524 CRC64;
SQ SEQUENCE

Query Match 6.1%; Score 94.5; DB 1; Length 446;
Best Local Similarity 25.8%; Pred. No. 2.9;

RP SEQUENCE FROM N.A.
RX MEDLINE=92165815; PubMed=1371507;
RA Cesak C., Martindale D.W.;
RT "Isolucyl-tRNA synthetase from the ciliated protozoan Tetrahymena
thermophila. DNA sequence, gene regulation, and leucine zipper
motifs."
RL J. Biol. Chem. 267:4592-4599(1992).
CC -1 CATALYTIC ACTIVITY: ATP + L-isoleucine + tRNA(Ile) = AMP +
CC diphosphate + L-isolucyl-tRNA(Ile).
CC -1 SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: M30942; AAA30122.1; -
CC PIR: A42399; A42399.
CC HSSP: P56690; 11LE.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR002301; tRNA-synt_1le.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00984; tRNA-synt_1le.
DR TIGRFAMs; TIGR00392; 1le; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_1; 1.
KW Aminoacyl-tRNA synthetase; protein biosynthesis; ligase; ATP-binding.
FT SITE 53 63 "HIGH" REGION.
FT SITE 607 611 "KMSKS" REGION.
FT BINDING 610 610 ATP (BY SIMILARITY).
SQ SEQUENCE 1081 AA; 124850 MW; D4D72616AFB8C795 CRC64;

Query Match 6.1%; Score 94; DB 1; Length 1081;
Best Local Similarity 19.2%; Pred. No. 8.9; Mismatches 103; Indels 140; Gaps 15;
Matches 67; Conservative 39;

QY 37 CVHFLANDNEIKKIOSVNLKEA-----PIKSAAE-----TFFSW----- 72
DB 190 CATVLN-----FETQONKYKEVDPSLFIAPKTAEDPKTFIAWTTPTWTLPSNLALVI 243
QY 73 -----YYKSKDG-----ECNELDKELKGGKIPPAF 98
DB 244 NKDFDYKVLDAKTOEHYLLAECLPELYKKKDKGYKILEKFGSELYGVEYEPLEPYFL 303
99 LKSMFTF-----GDYRDFLEGTDI---SKGHGSGSKLKEQIDSLFKNGDQKSP---NG-- 146
304 SRKODGCFRIILAGDFTADAGTGVHCAPGFGDDYKVSANNITIKPDDPPVPVDENGHF 363
QY 147 -----KTRQE-----WTEHSHEIWEAMCALVK 170
DB 364 TINVSDPAGVYIKKADKLKRNKLERGLLLVDSFKNYRPFQWRSDTPLYKAVHCFIK 423
QY 171 IGAKKDDFTENTGYNNVKSSTLEEFARKPQFLRMLEMTDYCYTRQY----- 223
DB 424 VFLAKDDLLAN-----NKKAYWVPKFAOEGREFNNMLQN-VSDMCFSSRFWGNPIPI 474
QY 224 --LKDVQF-----KCKSNQIAKODTECNKKCEQYVYKMKKKKKWIPQDK 265
DB 475 WSEDEFEVYVIGSVLELKKLTGATEITDLAKDFIDHILT-----IPSQK 518

Search completed: April 28, 2003, 10:30:26
Job time : 9.65193 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 10.3425 Seconds

(without alignments)
5498.552 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554

Perfect score: 1543
Sequence: 1 DCHPRKNSNGYPPDMQGNIN.....KKWIPDKYKDERDKRRF 276

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rv1rus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1543	100.0	3542	5	Q9U5M2
2	750	48.6	1615	5	Q9U5M2
3	745	48.3	3287	5	Q9U5M2
4	634.5	41.1	2706	5	Q9U5M2
5	624	40.4	3078	5	Q9U5M2
6	614	39.5	182	5	Q9U5M2
7	609	39.5	3026	5	Q9U5M2
8	557.5	36.1	2658	5	Q9U5M2
9	544	35.3	169	5	Q9U5M2
10	524	34.0	1685	5	Q9U5M2
11	516.5	33.5	2710	5	Q9U5M2
12	493.5	32.0	921	5	Q9U5M2
13	493	32.0	2527	5	Q9U5M2
14	486	31.5	3006	5	Q9U5M2
15	475	30.8	1711	5	Q9U5M2
16	453.5	29.4	195	5	Q9U5M2

17	447	29.0	176	5	Q9B0F0
18	439	28.5	184	5	Q9B0F0
19	410.5	26.6	2042	5	Q9B0F0
20	408.5	26.5	173	5	Q9B0F0
21	408.5	26.5	2209	5	Q9B0F0
22	402.5	26.1	2135	5	Q9B0F0
23	396	25.7	2212	5	Q9B0F0
24	390.5	25.3	2664	5	Q9B0F0
25	385.5	25.0	2664	5	Q9B0F0
26	383	24.8	2209	5	Q9B0F0
27	382.5	24.8	2182	5	Q9B0F0
28	371.5	24.1	2647	5	Q9B0F0
29	369.5	23.9	2924	5	Q9B0F0
30	363.5	23.6	2163	5	Q9B0F0
31	353	22.9	2277	5	Q9B0F0
32	350	22.7	162	5	Q9B0F0
33	347	22.5	160	5	Q9B0F0
34	343.5	22.3	177	5	Q9B0F0
35	341	22.1	2169	5	Q9B0F0
36	339.5	22.0	2197	5	Q9B0F0
37	299.5	19.4	1729	5	Q9B0F0
38	289.5	18.8	616	5	Q9B0F0
39	289.5	18.8	616	5	Q9B0F0
40	289.5	18.8	1210	5	Q9B0F0
41	289.5	18.8	1210	5	Q9B0F0
42	274	17.8	616	5	Q9B0F0
43	274	17.8	1475	5	Q9B0F0
44	270	17.5	616	5	Q9B0F0
45	269	17.4	616	5	Q9B0F0

ALIGNMENTS

RESULT 1

Q9U5M2 ID Q9U5M2 PRELIMINARY; PRT; 3542 AA.
AC Q9U5M2; 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2002 (TREMBLrel. 20, Last annotation update)
DE FCR3 CSA ligand (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MEDLINE=20006305; PubMed=10535993;
RA Buffet P., Gamain B., Scheidig C., Baruch B., Oishi S., Fujii N.,
RT "Plasmodium falciparum domain mediating adhesion to chondroitin
sulfate A: A receptor for human placental infection.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:12743-12748(1999).
DR EMBL; AJ13811; CAB59840.1;
DR InterPro: IPR001219; Neurotoxin.
DR Pfam: PF004258; PEMP.
DR PRINTS; PR00284; TOXIN.
FT NON_TER 3542 3542
SQ SEQUENCE 3542 AA; 413089 MW; 970D85EE88DA2EC2 CRC64;

Query Match 100.0%; Score 1543; DB 5; Length 3542;
Best Local Similarity 100.0%; Pred. No. 6.7e-109;

Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DCHPRKNSNGYPPDMQGNINLVEDPRVCMPRRQKLCVFLANDNEIKLQSOVNLKEAF 60
DB 1279 DCHPRKNSNGYPPDMQGNINLVEDPRVCMPRRQKLCVFLANDNEIKLQSOVNLKEAF 1338
QY 61 IKAAAEFFSWYKSKDEGNEIDKEKGIIPPAFLSMYTFEDYDFGFDISK 120

Db 1339 IKSAAETFFSWYKKKDGEGNELKELKGIIPAFLSMEYTFGDRDPLFGTDISK 1398
 Oy 121 GHGSGSKLRQIDSLFKNGDQKSPNGKTRQEWMTESHEIWEAMLCALVIGAKKDDFTE 180
 Db 1399 GHGSGSKLRQIDSLFKNGDQKSPNGKTRQEWMTESHEIWEAMLCALVIGAKKDDFTE 1458
 Oy 181 NYGNNVKKFSDKSTTLEEFARROPFLWLTWYDDCYTRQKYLKDVOEKCSNDOLKCD 240
 Db 1459 NYGNNVKKFSDKSTTLEEFARROPFLWLTWYDDCYTRQKYLKDVOEKCSNDOLKCD 1518
 Oy 241 TECNKKCEDYVKKKKKKEMIPDDKYKXKDERDKRRF 276
 Db 1519 TECNKKCEDYVKKKKKKEMIPDDKYKXKDERDKRRF 1554

RESULT 2

087325
 ID 087325 PRELIMINARY; PRT; 1615 AA.
 AC 087325;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DE PfEMP1 (Fragment).
 GN TM284VAR3.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TM284;
 RX MEDLINE=21927235; PubMed=11930336;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT "Identification of a conserved Plasmodium falciparum var gene
 RT implicated in malaria in pregnancy."
 RL J. Infect. Dis. 185:1207-1211(2002).
 DR EMBL; AJ420412; CAD20867.1; -
 FT NON_TER 1615
 SQ SEQUENCE 1615 AA; 188639 MW; FCA896C008B6DEA6 CRC64;

Query Match

Best Local Similarity 48.6%; Score 750; DB 5; Length 1615;
 Best Local Similarity 51.0%; Pred. No. 8.8e-49;
 Matches 150; Conservative 33; Mismatches 79; Indels 32; Gaps 9;

Oy 1 DCHPRKNSNGYPMOC---GNINLVEDPRVCMPPRRQKLCVHPLANDNEIKLQSOVNL 56
 Db 1262 ECNVATTEGE-YPEWCEKQIONIH---KGACMPRRRLCVHFLKE---LKVEIYEKL 1312
 Oy 57 KEAFIKSAAAEFFSWYKKSKDGGNELDKELKGIIPAFLSMEYTFGDRDPLFGT 116
 Db 1313 KEAFIQCASAEFFLLMKRYKEDNNGCEDLQNLGSIIPDDFKQMFYTFGDRDPLCGK 1372
 Oy 117 DISKGHGSGSKLRQIDSLFKNGDQKSPNGKTRQEWMTESHEIWEAMLCAL----- 168
 Db 1373 DIGSDVY---NNOKINLALRIQSDSD--KRNMMNGIKEDVWKGAVCGLSHDVSGNH 1427
 Oy 169 ---VATGAKKDDFENNGYNNVKSFSK--STTLEEFARROPFLWLTWYDDCYTRQY 223
 Db 1438 KEIYARKIMEDPQNNKYNSVKTDEPSGKTLSEFAKVPFLWLTWYDDCYTRQY 1487
 Oy 224 LKDVQEKCSNDOLKCDTECNKKCEDYVKKK-KKEWIPDDKYKXKDERDKRRF 276
 Db 1488 LKEVESTCKSNDOCLKDTECNKKCEDYVKKKKEWIPDDKYKXKDERDKRRF 1541

RESULT 3

087326
 ID 087326 PRELIMINARY; PRT; 3287 AA.
 AC 087326;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DE PfEMP1 (Fragment).
 GN TM180VAR2.

OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TM180;
 RX MEDLINE=21927235; PubMed=11930336;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT "Identification of a conserved Plasmodium falciparum var gene
 RT implicated in malaria in pregnancy."
 RL J. Infect. Dis. 185:1207-1211(2002).
 DR EMBL; AJ420411; CAD20867.1; -
 FT NON_TER 3287
 SQ SEQUENCE 3287 AA; 383550 MW; 58F9F866FC244536 CRC64;

Query Match

Best Local Similarity 48.3%; Score 745; DB 5; Length 3287;
 Best Local Similarity 50.7%; Pred. No. 4.7e-48;
 Matches 149; Conservative 34; Mismatches 79; Indels 32; Gaps 9;

Oy 1 DCHPRKNSNGYPMOC---GNINLVEDPRVCMPPRRQKLCVHPLANDNEIKLQSOVNL 56
 Db 1262 ECNVATTEGE-YPEWCEKQIONIH---KGACMPRRRLCVHFLKE---LKVEIYEKL 1312
 Oy 57 KEAFIKSAAAEFFSWYKKSKDGGNELDKELKGIIPAFLSMEYTFGDRDPLFGT 116
 Db 1313 KEAFIQCASAEFFLLMKRYKEDNNGCEDLQNLGSIIPDDFKQMFYTFGDRDPLCGK 1372
 Oy 117 DISKGHGSGSKLRQIDSLFKNGDQKSPNGKTRQEWMTESHEIWEAMLCAL----- 168
 Db 1373 DIGSDVY---NNOKINLALRIQSDSD--KRNMMNGIKEDVWKGAVCGLSHDVSGNH 1427
 Oy 169 ---VATGAKKDDFENNGYNNVKSFSK--STTLEEFARROPFLWLTWYDDCYTRQY 223
 Db 1428 KEIYARKIMEDPQNNKYNSVKTDEPSGKTLSEFAKVPFLWLTWYDDCYTRQY 1487
 Oy 224 LKDVQEKCSNDOLKCDTECNKKCEDYVKKK-KKEWIPDDKYKXKDERDKRRF 276
 Db 1488 LKEVESTCKSNDOCLKDTECNKKCEDYVKKKKEWIPDDKYKXKDERDKRRF 1541

RESULT 4

015870
 ID 015870 PRELIMINARY; PRT; 2706 AA.
 AC 015870;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DE PfEMP1 (Fragment).
 GN R29R4VAR1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IT 4/25/5;
 RX MEDLINE=97373957; PubMed=9230440;
 RA Rowe J.A., Moulds J.M., Newbold C.I., Miller L.H.;
 RT "Plasmodium falciparum rosetting is mediated by PfEMP1 and requires
 RT complement receptor 1."
 RL Nature 388:292-295(1997).
 DR EMBL; Y13402; CAA73831.1; -
 DR EMBL; Y13403; CAA73831.1; JOINED.
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR InterPro: IPR004258; PfEMP.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 FT NON_TER 2706
 SQ SEQUENCE 2706 AA; 308162 MW; 1C33D55AD5317D68 CRC64;

Query Match

Best Local Similarity 41.1%; Score 634.5; DB 5; Length 2706;
 Best Local Similarity 45.5%; Pred. No. 1e-39;
 Matches 131; Conservative 37; Mismatches 91; Indels 29; Gaps 10;

QY 1 DCHPRKNSNGYDPDQGNINIVE-DRVCGMPRRKOLCVHFLANDNEIKKLOSQVNIKEA 59
DB 787 NCAAKRRKN--EMQCDKMTFVGNNEGVCMPRRKRSICIHNLLEQOTK---NKYOLREA 841
QY 60 FIKSAAEFFSWYYSKSGKDEGNELDKELKEGKIPPAFLRSFYEGYDRDLFGDIS 119
DB 842 FIKCAKENTLMDKKNKNEAEL---LKKKIPEDPRIMFYTFGDFRDLCLDNG 898
QY 120 KGHGEGSKLEQIDSLFKNGDOKSPNGKTRQEWTEHSHIEWAMCALVKGAK----- 174
DB 899 K--DVDKYNKINIKYFNNSKRGFKKIDPENWMNNGQIWMGMCALHADTKSIRK 955
QY 175 KODFTNYGYNVNFSDK-----STLEBPAKPOFLRMTEYDYCYTRKQYLDVQE 229
DB 956 KD---NYREKTYLAKRDGSGMGLSEPAKPKFLRFVEYDDYCKERKQYLTVEAS 1011
QY 230 KCKSND--OLKCDTECNKCEYVYKMKKK--EMIPODKYYKDERDK 274
DB 1012 TCKSIDGGLKCDRGCKNCKDEYKTKMKKKKEEMNLQDKYTKDEREK 1059

RESULT 5

Q26031 PRELIMINARY; PRT: 3078 AA.
AC 026031:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Variant-specific surface protein.
GN VAR-1.
OS Plasmodium falciparum (isolate Dd2).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=57267;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DD2;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Gulnet F., Herrfeldt J.A., Peterson D.S., Ravetch J.A., Wellens T.E.;
RT "The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum- infected erythrocytes."
RT Cell 82:89-100(1995).
RL EMBL: LA0608; AAA75396.1;
DR InterPro: IPR004258; PfEMP.
DR Pfam: PF03011; PfEMP. 2.
SQ SEQUENCE 3078 AA; 349297 MW; C8037C2BC3CCD7C3 CRC64;

Query Match

40.4%; Score 624; DB 5; Length 3078;

Best Local Similarity 43.8%; Pred. No. 7.5e-39; Mismatches 83; Indels 32; Gaps 11;

QY 1 DCHPRKNSNGYDPDQGNINIVE-DRVCGMPRRKOLCVHFLANDNEIKKLOSQVNIKEA 59
DB 1365 ECHPKRS---YPMDCKNNDIDSH-DGACMPFRKOLCLYIHESQTEIKTDNLIKDA 1420
QY 60 FIKSAAEFFSWYYSKSGKDEGNELDKELKEGKIPPAFLRSFYEGYDRDLFGDIS 118
DB 1421 FIKTAAEFFLSQWYYSKSKNDSEAKILDR---GLIPSGFLRSMYTFGDRIDICMTDI 1476
QY 119 SKHGEGSKLEQIDSLFKNGDOKSPNGKTRQEWTEHSHIEWAMCALVKGAKKD- 177
DB 1477 SKQONVNAKAKIDIGKFFSKDGSPPSGLSRDMWMTKNGEIKMGMCALTKVTTDTDK 1536
QY 178 --FTENYGNVNFSDK--STLEBPAKPOFLRMTEYDYCYTRKQYLDVQEKCS- 233
DB 1537 RKIKNDYSTDVQNGNNGNLEEFPAKPOFLRMTEYDYCYTRKQYLDVQEKCS- 1596
QY 234 NDOLKCDT---ECNKKCEDYVYKMK--KKKEWI-----PDQKYYK 268
DB 1597 NSTQOCNDAAKRCNOACRAVQOEYVENKKKEFSQGTNNFVLKANVQPDPEYK 1648

RESULT 6

ID 000832 PRELIMINARY; PRT: 182 AA.
AC 000832:
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Partial erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum (isolate Dd2).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=57267;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DD2;
RX MEDLINE=97342604; PubMed=9199301;
RA Fischer K., Horrocks P., Preuss M., Wiesner J., Wensch S., Camargo A.A., Lanzer M.;
RT "Expression of var genes located within polymorphic subtelomeric domains of Plasmodium falciparum chromosomes."
RT Mol. Cell. Biol. 17:3679-3686(1997).
DR EMBL: Y11910; CAA72661.1;
FT NON_TER 1
FT SEQUENCE 182 AA; 20830 MW; 197C4A677BAD148B CRC64;

Query Match

39.8%; Score 614; DB 5; Length 182;

Best Local Similarity 63.7%; Pred. No. 1.7e-39; Mismatches 34; Indels 4; Gaps 4;

QY 31 PROKLCVHFLANDNEIKKLOSQVNIKEAFIKSAAEFFSWYYSKSGDE-GNELDKEL 89
DB 1 PROKLCVHFLANDNEIKKLOSQVNIKEAFIKSAAEFFSWYYSKSGDE-GNELDKEL 60
QY 90 KEKIPPAFLRSFYEGYDRDLFGDISKSGHGEKSLKEQIDSLF-KRGDDKSPNGKT 148
DB 61 KKGITPEFLRSFYEGYDRDLFGDISKSGHGEKSLKEQIDSLF-KRGDDKSPNGKT 120
QY 149 ROEWTEHSHIEWAMCALVKI-GAKKDFTEYGNVNFSDK--STLEBPAKPOFL 206
DB 121 REDWMTKNGPDWKMGMICALTNASGAKKETLTDKYYKESVTFGDSGPNLQTFSSRQFL 180
QY 207 RW 208
DB 181 RW 182
RESULT 7
Q26030 PRELIMINARY; PRT: 3026 AA.
AC 026030:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 08, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Variant surface protein (fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE OF 2044-2922 FROM N.A.
RC STRAIN-IT 4/25/5;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Gulnet F., Herrfeldt J.A., Peterson D.S., Ravetch J.A., Wellens T.E.;
RT "The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum- infected erythrocytes."
RT Cell 82:89-100(1995).
RL [1]
RN [1]
RP SEQUENCE OF 2044-2922 FROM N.A.
RC STRAIN-IT 4/25/5;
RA Smith J.D., Chitnis C.E., Craig A.G., Roberts D.J.,

Chatropadhyay R., Pillai C.R., Chitnis C.;
"Identification of a domain responsible for binding to intercellular

121 NGPDIWKMCLALTNASGAKKETLTLDKYKYESTYFGDNGSPNLOTFFSSR 169

```

RESULT 10
ID 0904A2 PRELIMINARY; PRT; 1685 AA.
AC 0904A2;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAY-2002 (TREMblrel. 20, Last annotation update)
DE Variant surface protein PfEMP1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IT;
RA MEDLINE=20144115; PubMed=10677532;
RA Smith J.D., Craig A.C., Krick N., Hudson-Taylor D., Kyes S., Fagen T.,
RA Pinches R., Baruch D.I., Newbold C.I., Miller L.H.;
RT Identification of a Plasmodium falciparum intercellular adhesion
RT molecule-1 binding domain: A parasite adhesion trait implicated in
RT cerebral malaria.
RA Proc. Natl. Acad. Sci. U.S.A. 97:1766-1771(2000).
DR EMBL: AF193424; AAF1980.1; -.
DR InterPro: IPR000454; Eub_Atpase_Gsub.
DR InterPro: IPR004258; PfEMP.
DR Pfam: PF03011; PfEMP; 1.
DR PROSITE: PS00605; ATPase_C; 1.
FT NON TER 1685 1685
SQ SEQUENCE 1685 AA; 193671 MW; D1FDA266680551E CRC64;

Query Match 34.0%; Score 524; DB 5; Length 1685;
Best Local Similarity 42.7%; Pred. No. 1.6e-31;
Matches 122; Conservative 35; Mismatches 89; Indels 40; Gaps 11;

QY 14 WCGNINLVDPVPCMPRRROKLCVHFLANDNEIKKIQSQVNLKEAFIKSAAEFFSWY 73
DB 1252 WKCENTKLGDEGCMPPRRONLCVHLYTKLNDKSK--EEDLEAFIKSAAEFFLLRQ 1308
QY 74 YKSKDGEKNELDKELKEKIPPAFLRSMTYTFGDRDFGTDISK--SHGEGSKLKEQ 131
DB 1309 YNYSKNVED--DKILHRDMIPPEFRSMYTFGDRYDIDCLDIDISKIADHVTYAKK 1365
QY 132 ISLFFKNGDOKSPNGK---TRQEMWTEHSHEIWEAMCAL-VKIGAKKDD----FTENY 182
DB 1366 ITAVFOKIGSKTNGKVTLEREGMKVEYGLSIWGMICALSYMTETKMKDEGVYTYLMKY 1425
QY 183 GYNNVAFSDKSTLEEFARPOFLRMTEWYDYCYTRQKYLNDVOKKCS-----ND 235
DB 1426 IYKN---NDIKYLEEPASPPPLRMWTEGDFVKNRKKEVLVLAKKCDSCITLRNNGTS 1482
QY 236 QLKCDPE-----CNKCEDYVKKMKKKKEWIPQDKYKDRDKRRF 276
DB 1483 NKTKCDNENGACKTQCEKKKKMERWK-----KHTSSQ--KKKF 1520

RESULT 11
ID 09XZB8 PRELIMINARY; PRT; 2710 AA.
AC 09XZB8;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Variant-specific surface protein.
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS2;
RA MEDLINE=99238507; PubMed=10220443;
RA Reeder J.C., Cowman A.F., Davern K.M., Beeson J.G., Thompson J.K.,
RA Rogerson J.J., Brown G.V.;

```

```

RT "The adhesion of Plasmodium falciparum-infected erythrocytes to
RT chondroitin sulfate A is mediated by P. falciparum erythrocyte
RT membrane protein 1."
RL Proc. Natl. Acad. Sci. U.S.A. 96:5198-5202(1999).
DR EMBL: AF134154; AAD29126.1; -.
DR InterPro: IPR004258; PfEMP.
DR Pfam: PF03011; PfEMP; 2.
SQ SEQUENCE 2710 AA; 309247 MW; 0AC64F0A3D5BF512 CRC64;

Query Match 33.5%; Score 516.5; DB 5; Length 2710;
Best Local Similarity 40.4%; Pred. No. 1e-30;
Matches 118; Conservative 43; Mismatches 100; Indels 31; Gaps 12;

QY 2 CHPKNSNGYPRDQCQGNINLVDPVPCMPRRROKLCVHFLANDNEIKKIQSQVNLKEAFI 61
DB 918 CNRKES---YPRDQCKKNIDNSHSGACMPRRROKLCVRLDLOGEIKRPE---DILKFI 971
QY 62 KSAAEFFFSWYKSKDGEKNELDKELKEKIPPAFLRSMTYTFGDRDFLEGTDIS-- 119
DB 972 NCAKETHFAMHMYK-KDYNNAE--NELKSGKIPGFRKOMYTFGDRDFLEGTDISSC 1028
QY 120 ---KGHEGSKLKEQIDSLFKNGDOKSPNGKTRQEMWTEHSHEIWEAMCALVK--IGAK 174
DB 1029 RYIKDVSQITKSLGQDATTKEGDTHTDNNKKEDEMTIHGPKIWEGLCALTNGLSESE 1088
QY 175 KDFETENYGYNNVAFSDK-STLEEFARPOFLRMTEWYDYCYTRQKY---LKDY--- 227
DB 1089 KKNILDYSYNNKLNNAKDKDCCLEKFPASKPOLFLMWYEMSEDFEREKKKLEADVEDCIK 1148
QY 228 ---QEKSKSDQKCDTECNKCKEDYVYK-KKEWIPD-KYKDERDK 274
DB 1149 AKDYGCKNN---KSNNSCVKVEYENYITGKTYQESQEGKNTKROKK 1197

RESULT 12
ID 025989 PRELIMINARY; PRT; 921 AA.
AC 025989;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE (clone pS31H) ORF (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DD2/NM;
RA MEDLINE=95350219; PubMed=7624377;
RA Peterson D.S., Miller L.H., Wellens T.E.;
RT "Isolation of multiple sequences from the Plasmodium falciparum genome
RT that encode conserved domains homologous to those in erythrocyte-
RT binding proteins."
RL Proc. Natl. Acad. Sci. U.S.A. 92:7100-7104(1995).
DR EMBL: L38454; AAC37240.1; -.
DR InterPro: IPR004258; PfEMP.
DR Pfam: PF03011; PfEMP; 1.
FT NON TER 1
SQ SEQUENCE 921 AA; 104711 MW; D98BC07763628746 CRC64;

Query Match 32.0%; Score 493.5; DB 5; Length 921;
Best Local Similarity 40.0%; Pred. No. 1.7e-29;
Matches 124; Conservative 39; Mismatches 94; Indels 53; Gaps 15;

QY 1 DCHPKNSNGYPRDQCQGNINLVDPVPCMPRRROKLCVHFLANDNEIKKIQSQVNLKEAFI 58
DB 414 EYRKET---YSEWTCDESKIKMGQHGACIPPRROKLCIHL-----EKIMTNTNELKY 464
QY 59 AFKISAAETFFSWY-YKSKDGEKNELDKELKEKIPPAFLRSMTYTFGDRDFLEGTDT 117
DB 465 AFKICAAETFFLMQNKKKKNGAEDLDEKLGITIPEDFKROMFTFPADYRICGTD 524
QY 118 I-----SKHGEGSKLKEQIDSLFKNGDOKSPNG-KTRQEMWTEHSHEIWEAMCAL-- 168

```

Db 525 ISSKDTSGVGV---KVCNIDVF-----YKISNIRKSKWMEINGVINEGHCALST 577
 QY 169 -----YIGAKKDDFTENYGNANKF--SDKSTLEBAKAPQILRWLTWYDYCYTRQ 221
 Db 578 DTSLNANNPEHTKLTETGNNNNEKVIYFGSDSTLSTKSESRPOFLRWLTENGNECKOK 637
 QY 222 KYLKVQCKSKNDQKCDTECNK--CEDYVK-----YMKKKKEMIPQDK- 265
 Db 638 KEYKVLAKCKDCD-VDGDGKNGCKVACKQCKQYHSWIGIMDNKKKOKGRTEVKKI 696
 QY 266 -YKDERDK 274
 Db 697 PLYKEDKVK 706

RESULT 13
 ID 095W83 PRELIMINARY; PRT; 2527 AA.

NC 01-DEC-2001 (Tremblrel. 19, Created)
 DE 01-DEC-2001 (Tremblrel. 19, last sequence update)
 DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
 GN Erythrocye membrane protein 1 (Fragment).
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN NCBL_TaxID=5833;
 RN [1]
 RX MEDLINE-21442075; PubMed-11557894;
 RA Flick K., Scholander C., Chen Q., Fernandez V., Pouvelle B., Gysin J.,
 RT "Role of nonimmune IgG bound to PfEMP1 in placental malaria."
 RL Science 293:2098-2100(2001).
 DR EMBL: AF366567; AAL12845.1;
 DR InterPro: IPR004258; PfEMP.
 DR Pfam: PF03011; PfEMP; 2.
 FT NON_TER 2527 2527
 SQ SEQUENCE 2527 AA; 292137 MW; 59E2766BF0425E4 CRC64;

Query Match Best Local Similarity 32.0%; Score 493; DB 5; Length 2527;
 Matches 116; Conservative 50; Mismatches 95; Indels 76; Gaps 12;

QY 1 DCHPKKNSG---YDPDQGNINL---VEDPRVCMPPRQKLCVFLANDNEIKKLOSOV 54
 Db 1663 DACEEYKNGKTAAYPGMKCTTKIEKGEKEDVYVCIIPRQKLYI-----KQLEFSGKK 1716
 Db 55 N--LKEAFKSAAAETFSWYI-----KSKDGEKNELDKEL 89
 Db 1717 NDELKRAFLTECAAVETFEFAMHKYKEDKNEKEETSYLVRLGFLGKOSSPDEAOKDL 1776
 QY 90 KEGKIPPAFLRMETFGYRDLFGTIDISKHGEGR--LEQIDSLFRKNGDQSPNGKT 148
 Db 1777 ESKGIPDEKRRMFTYFGYRDLFGF---KNMGIDMTVEKIKSIPP--SVKPDQOT 1830
 QY 149 ROEWMTESHEITWMLCAL-----YIGAKKDDFTENYGNANKFSTLEBAK 201
 Db 1831 RETWNNNAKDIWNGMICALSYDTKRIKIGIYQNLQNDNDKYVITDNNTLEIYAT 1890
 QY 202 RPOFLRWLTWYDYCYTRQKYLKDVQCKSK-----NDOLKCD- 240
 Db 1891 VPOFLRSEEAWEFCRRKRLKHLAIHKDCRSDFPGMKYCSGDGHDCTNEELKINDPAD 1950
 QY 241 ---TECNKCEDYKVKKK--KEMIPQDKYKDERDK 273
 Db 1951 SYCPDCKKACRKYKWKIKEYEFTYKSKYKKEHLK 1987

RESULT 14
 ID 026032 PRELIMINARY; PRT; 3006 AA.

AC Q26032;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE Variant-specific surface protein.
 GN VAR-3.

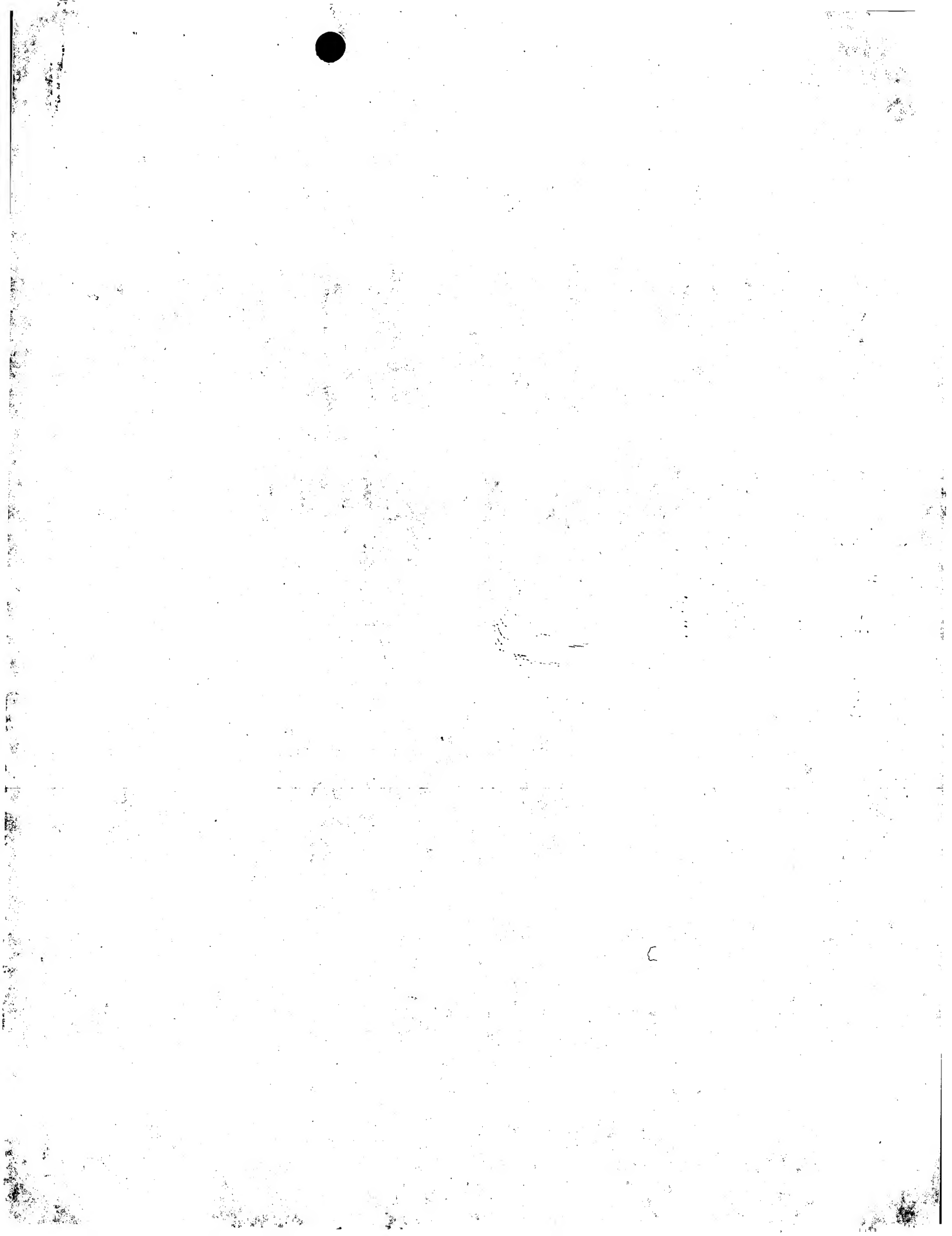
OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN NCBL_TaxID=5833;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=FCR3;
 RX MEDLINE-95330813; PubMed-7606788;
 RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
 RA Peterson D.S., Ravetch J.A., Wellem T.E.;
 RT "The large diverse gene family var encodes proteins involved in
 cytoadherence and antigenic variation of Plasmodium falciparum-
 infected erythrocytes."
 RL Cell 82:89-100(1995).

DR EMBL: LA0609; AAK73397.1;
 DR InterPro: IPR004258; PfEMP.
 DR Pfam: PF03011; PfEMP; 2.
 SQ SEQUENCE 3006 AA; 343769 MW; 8FD5FF475F943C74 CRC64;

Query Match Best Local Similarity 31.5%; Score 486; DB 5; Length 3006;
 Matches 115; Conservative 39; Mismatches 108; Indels 32; Gaps 13;

QY 5 KNSNGYDPMOCGNINLVEDPRVCMPPRQKLCVFLANDNEIKKLOSOVNLKRAFKSA 64
 Db 1247 KAGKDYKGMPCNSQIHHTHNGACMPPRQKLCVSGLTDRITALE---YIRIEFKSA 1303
 QY 65 AAEFTFSWYKSKDGEENELDKLECKGIPPAFLRMETFGYRDLFGTIDISKHGE 124
 Db 1304 ALEHFAWDRYKEDGEA---EAEIKNGNIPEGFRQWYTFGGRDIFPGROIST-HAY 1359
 QY 125 GSKLEQIDSLF-KNGDQK-----SPNGKTRQEWMTESHEITWMLCALV-KIG--AKK 175
 Db 1360 ISGVSPVITILEKENAKYAKONSNNELDDWDHGDHGMICALTHKISDEKK 1419
 QY 176 DDFTEINGYNNVKSCK-STTLEBPARRPOFLRWLTWYDYCYTRQKYLKDVQCK- 231
 Db 1420 KEINKKSYKLLNSPQKSNVDEFAKPPFLRFIEWGEDEFCQOREKRAKYKSSDA 1479
 QY 232 KSNOLKCDTECNK-----EDYKVKMKKKKEMIPQDKYKDER--DKKRF 276
 Db 1480 KDYDCK-NIKSNASCVSACKVYEDI--TKKVEYTKQKGFDAEKIDTKEGY 1530

RESULT 15
 ID 096108 PRELIMINARY; PRT; 1711 AA.
 AC 096108;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE PfEMP1.
 GN PFB0010W.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN NCBL_TaxID=5833;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE-99021743; PubMed-9804551;
 RA Gardner M.J., Tetteil H., Carucci D.J., Cummings L.M., Aravind L.,
 RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,
 RA Shen K., Jing J., Aston C., Mason T., Yu K., Fujii C., Pederson J.,
 RA Salzberg S., Zhou L., Sutton G.G., Schwartz D.C., Pertea M.,
 RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.,
 RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
 falciparum."
 RL Science 282:1126-1132(1998).
 DR EMBL: AE001366; AAC71792.1;



GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 2.78453 Seconds
(without alignments)
2916.372 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554

Perfect score: 1543
Sequence: 1 DHPKNSNGIPDMOCNIN.....KKWIPDKYDERDKRRF 276

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

tal number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

1: /cgcn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgcn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgcn2_6/prodata/1/1aa/5A.COMB.pep:*
4: /cgcn2_6/prodata/1/1aa/5B.COMB.pep:*
5: /cgcn2_6/prodata/1/1aa/5A.COMB.pep:*
6: /cgcn2_6/prodata/1/1aa/5B.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	624	40.4	2710	2	US-08-568-459A-12
2	624	40.4	2710	2	US-08-487-826B-12
3	624	40.4	2710	4	US-09-210-288-12
4	624	40.4	3060	2	US-08-487-826B-14
5	493.5	32.0	921	2	US-08-568-459A-8
6	493.5	32.0	921	2	US-08-487-826B-8
7	493.5	32.0	921	4	US-09-210-288-8
8	382.5	24.8	2182	2	US-08-487-826B-16
9	274	17.8	1435	2	US-08-568-459A-4
10	274	17.8	1435	2	US-08-487-826B-4
11	274	17.8	1435	4	US-09-210-288-4
12	273.5	17.7	311	2	US-08-568-459A-21
13	273.5	17.7	311	2	US-08-487-826B-33
14	273.5	17.7	311	4	US-09-210-288-21
15	260	16.9	749	2	US-08-568-459A-6
16	260	16.9	749	2	US-08-487-826B-6
17	260	16.9	749	4	US-09-210-288-6
18	223	14.5	1115	2	US-08-568-459A-2
19	223	14.5	1115	2	US-08-487-826B-2
20	223	14.5	1115	4	US-09-210-288-2
21	223	14.5	1115	6	5198347-6
22	218.5	14.2	324	2	US-08-568-459A-17
23	218.5	14.2	324	2	US-08-487-826B-29
24	218.5	14.2	324	4	US-09-210-288-17
25	210	13.6	700	2	US-08-568-459A-10
26	210	13.6	700	2	US-08-487-826B-10
27	210	13.6	700	4	US-09-210-288-10

28	172.5	11.2	197	6	5198347-2	Patent No. 5198347
29	172.5	11.2	778	6	5198347-4	Patent No. 5198347
30	164	10.6	277	2	US-08-568-459A-15	Sequence 15, Appl
31	164	10.6	277	4	US-08-487-826B-27	Sequence 27, Appl
32	164	10.6	277	2	US-09-210-288-15	Sequence 15, Appl
33	164	10.6	277	2	US-08-568-459A-16	Sequence 16, Appl
34	142	9.2	282	2	US-08-487-826B-28	Sequence 28, Appl
35	142	9.2	282	4	US-09-210-288-16	Sequence 16, Appl
36	140	9.1	291	2	US-08-568-459A-13	Sequence 13, Appl
37	140	9.1	291	2	US-08-487-826B-25	Sequence 25, Appl
38	140	9.1	291	4	US-09-210-288-13	Sequence 13, Appl
39	133.5	8.7	411	2	US-08-568-459A-19	Sequence 19, Appl
40	133.5	8.7	411	2	US-08-487-826B-31	Sequence 31, Appl
41	133.5	8.7	411	4	US-09-210-288-19	Sequence 19, Appl
42	126	8.2	411	2	US-08-568-459A-20	Sequence 20, Appl
43	126	8.2	411	2	US-08-487-826B-32	Sequence 32, Appl
44	126	8.2	411	4	US-09-210-288-20	Sequence 20, Appl
45	123	8.0	362	2	US-08-568-459A-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-08-568-459A-12
Sequence 12, Application US/08568459A

Patent No. 5849306

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OR INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/568,459A

FILING DATE: 07-DEC-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Israelsen, Ned

REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH121.001CP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ORGANISM: Plasmodium falciparum

Query Match 40.4% Score 624 DB 2 Length 2710

Best Local Similarity 43.8%, Pred. No. 8.5e-53;
Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11;

```

QY 1 DCHPKNSNGYEDMOC--NININVEDPRVCMPPRRQKLCVHLANDNEIKKIOSVNLKEA 59
DB 1367 ECNPKES---YDMDCKNNIDISHD-GACMPRRQKLCVYIAHSQTEENIKTDNLDA 1422
QY 60 FIKSAAEFFESWYKSK--DGEENELDKELKGRIPAPLRSMTYTGDRDFLFGDI 118
DB 1423 FIKTAAEFFLSWQYKSKNSSEAKILDR---GLIPQFLRSMTYTGDRDFLFGDI 1478
QY 119 SKGHGSKLAKQIDSLFNGQKSPNGKTRQEWMTESHSHEIWEAMLCALYKIGAKKD- 177
DB 1479 SKQNDVAAKADKIGKFFSKDSKSPGSLRQEWMTKNGPEIWMKMLCALYKVTDTDNK 1538
QY 178 --FTENYGNVNFSDK--STLLEFAKRPQFLRWLTJEMYDDCYTROYKYLKDVQEKCS- 233
DB 1539 RKINDYSYDKVNOSONGNSPSEFAKRPQFLRWMTJEMGEFCAEROKKENIINDACNEI 1598
QY 234 NDOLKCDT---ECNKKCEDYKYM-KKKEWI-----PODKYYK 268
DB 1599 NSTQOCNDAKRCNQACRAVOEYVENKKKEFSGQTNNEVLKANYQPDPEYK 1650

```

RESULT 2

US-08-487-8268-12
Sequence 12, Application US/084878268
Patent No. 5993827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,8268
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-487-8268-12

Query Match

Best Local Similarity 40.4%; Score 624; DB 2; Length 2710;
Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11;

```

QY 1 DCHPKNSNGYEDMOC--NININVEDPRVCMPPRRQKLCVHLANDNEIKKIOSVNLKEA 59
DB 1367 ECNPKES---YDMDCKNNIDISHD-GACMPRRQKLCVYIAHSQTEENIKTDNLDA 1422
QY 60 FIKSAAEFFESWYKSK--DGEENELDKELKGRIPAPLRSMTYTGDRDFLFGDI 118
DB 1423 FIKTAAEFFLSWQYKSKNSSEAKILDR---GLIPQFLRSMTYTGDRDFLFGDI 1478
QY 119 SKGHGSKLAKQIDSLFNGQKSPNGKTRQEWMTESHSHEIWEAMLCALYKIGAKKD- 177
DB 1479 SKQNDVAAKADKIGKFFSKDSKSPGSLRQEWMTKNGPEIWMKMLCALYKVTDTDNK 1538
QY 178 --FTENYGNVNFSDK--STLLEFAKRPQFLRWLTJEMYDDCYTROYKYLKDVQEKCS- 233
DB 1539 RKINDYSYDKVNOSONGNSPSEFAKRPQFLRWMTJEMGEFCAEROKKENIINDACNEI 1598
QY 234 NDOLKCDT---ECNKKCEDYKYM-KKKEWI-----PODKYYK 268
DB 1599 NSTQOCNDAKRCNQACRAVOEYVENKKKEFSGQTNNEVLKANYQPDPEYK 1650

```

RESULT 3

US-09-210-288-12
Sequence 12, Application US/09210288
Patent No. 6392026

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-12

Query Match	40.4%;	Score 624;	DB 4;	Length 2710;
Best Local Similarity	43.8%;	Pred. NO. 8.5e-53;		
Matches 128; Conservative	49;	Mismatches 83;	Indels 32;	Gaps 11

QY	1	DCHRRKSNQYPMOC-GNINIVYEDRVCVMPROKCYVHFLANDNEIKKLOSQYNIKEA	59
Db	1367	ECANKEK---YPMODCKNNNDISHD-GACAPPRKQICLYTAHSHQSYENIKITDNDNLDA	14222
QY	60	FIKSAAAETFEFSWYYKSR-DGEGNBLDELEKGIPIPAFLSMSEYTGEDYNDLFEGTDI	118
Db	1423	FIKFAAAETFEFSWYYKSNKSDSEAKIILDR---GLIISOFLSMSEYTGEDYNDIILANDI	1478
QY	119	SKHGSESKLKEBOLDLFFKNQDQSPNGKTRQBMWHEHSHETWEMLCALYIAGKKRD-	177
Db	1479	SKKQNDYAAKKKDKIKGFFESDQSGSSGSLSRQMMWTTNGPEIWKGLICLYITVYTDIYNK	1538
QY	178	--PFTENYGNVNVKFSDK-STTLEFPAKRPQFLMILTEWYDLCYTRQKYLKDVOECKRS-	233
Db	1539	RKIKNDYSYDKVNOVONGNSLEEFPAKQFOLTEWMIEMGEFCAEROKKENIKRACNEI	1598
QY	234	NDOLKQCT---ECKKCEDEDVYKTK-KKKEMI-----PODKYK 268	
Db	1599	NSTOOCCHDAHRCNQACRAVOEYVENKKKEFSQTNFNLPAKANQVPODEYK 1650	

RESULT 4

US-08-487-826B-14
; Sequence 14, Application US/08487826B
; Patent No. 5993827

GENERAL INFORMATION:

1 APPLICANT: Sim, Kim L.
 2 APPLICANT: Chluis, Chetan
 3 APPLICANT: Miller, Louis H.
 4 APPLICANT: Peterson, David S.
 5 APPLICANT: Su, Kin-zhaun
 6 APPLICANT: Wellens, Thomas E.
 7 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 8 TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 9 NUMBER OF SEQUENCES: 45
 10 CORRESPONDENCE ADDRESS:
 11 ADDRESSEE: Knobbe Martens Olson & Bear
 12 STREET: 620 Newport Center Drive 16th Floor
 13 CITY: Newport Beach
 14 STATE: California

• COMPUTER READABLE FORM:

1 MEDIUM TYPE: Floppy disk
2 COMPUTER: IBM PC compatible
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: PatentIn Release #1.0, Version #1.25
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/08/487,826B
7 FILING DATE: 10-SEP-1993

Query Match 40.48; Score 624; DB 2; Length 3060;

Best Local Similarity 43.8%; Pred. NO. 1e-52;
Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11.

[illegible]

RESULT 5

US-08-568-459A-8
; Sequence 8, Application US/08568459A
; Patent No. 5849306

; GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chltnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California

COMPUTER READABLE FORM:

1
 2 MEDIUM TYPE: Floppy disk
 3 COMPUTER: IBM PC compatible
 4 OPERATING SYSTEM: PC-DOS/MS-DOS
 5 SOFTWARE: Patent Release #1.0, Version #1.25
 6
 7 CURRENT APPLICATION DATA:
 8
 9 APPLICATION NUMBER: US/08/568,459A
 10
 11 FILING DATE: 07-DEC-1995

Query Match 32.0%; Score 493.5; DB 2; Length 921;
Best Local Similarity 40.0%; Pred. No. 1.8e-40;
Matches 124; Conservative 39; Mismatches 94; Indels 53; Gaps 15;

OY 1 DCHPKNSNGYPMOCGINL-VEPRVCMPRRQKLCVHFLANDNEIKLOSQVN-LKE 58
DB 414 ECRKFT---YSEWTCDESKIKMGOGACIPPRQKLCIHL-----EKIMTNNEIKY 464
OY 59 AFKSAAEFFSWY-YKSKDGESELDKELKSGKIPPAFLRSMFYFGDYRDLFGTD 117
DB 465 AFKCAAEFTFLMWNKKKNGAEDDEKLGITIPEDFKROMFTFFADYRDLCLGTD 524
OY 118 I-----SKHGEGSKLKEQDLSFKNGDQSPNG-KTROEMTSHSEIWEAMICAL-- 168
DB 525 ISSKDTSGVG---KVCNIDVF---YKISNIRYRKSMWETNGPYIWEGMICALSY 577
OY 169 -----VKIGAKKDDFTENYGYNNVF-SDKSTLEEFARPOFLRWLEWYDYCYTRQ 221
DB 578 DTSLNVPNPEHKKLTGNNNEKVIIFGSDSSTILSKFSERPOLRWLTGEMFCKEOK 637
OY 222 KYLDVQEKCKSNOLKCDTECNK---CEDYK-----YMKKKKEWIPQDK- 265
DB 638 KEYVLAKCKDCD-VGDGCKNGCVACKDCKQYHSHWIGIWDNYKKQGRYTEVKKI 696
OY 266 -YYKDERDK 274
DB 697 PLYKEDKDYK 706

RESULT 6
US-08-487-826B-8
Sequence 8, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-0176
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-487-826B-8

Query Match 32.0%; Score 493.5; DB 2; Length 921;
Best Local Similarity 40.0%; Pred. No. 1.8e-40;
Matches 124; Conservative 39; Mismatches 94; Indels 53; Gaps 15;

OY 1 DCHPKNSNGYPMOCGINL-VEPRVCMPRRQKLCVHFLANDNEIKLOSQVN-LKE 58
DB 414 ECRKFT---YSEWTCDESKIKMGOGACIPPRQKLCIHL-----EKIMTNNEIKY 464
OY 59 AFKSAAEFFSWY-YKSKDGESELDKELKSGKIPPAFLRSMFYFGDYRDLFGTD 117
DB 465 AFKCAAEFTFLMWNKKKNGAEDDEKLGITIPEDFKROMFTFFADYRDLCLGTD 524
OY 118 I-----SKHGEGSKLKEQDLSFKNGDQSPNG-KTROEMTSHSEIWEAMICAL-- 168
DB 525 ISSKDTSGVG---KVCNIDVF---YKISNIRYRKSMWETNGPYIWEGMICALSY 577
OY 169 -----VKIGAKKDDFTENYGYNNVF-SDKSTLEEFARPOFLRWLEWYDYCYTRQ 221
DB 578 DTSLNVPNPEHKKLTGNNNEKVIIFGSDSSTILSKFSERPOLRWLTGEMFCKEOK 637
OY 222 KYLDVQEKCKSNOLKCDTECNK---CEDYK-----YMKKKKEWIPQDK- 265
DB 638 KEYVLAKCKDCD-VGDGCKNGCVACKDCKQYHSHWIGIWDNYKKQGRYTEVKKI 696
OY 266 -YYKDERDK 274
DB 697 PLYKEDKDYK 706

RESULT 7
US-09-210-288-8
Sequence 8, Application US/09210288
Patent No. 6392026
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1PMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-0176
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:

```

?      LENGTH:   921 amino acids
?      TYPE:     amino acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: protein
?      HYPOTHETICAL: NO
?      ORIGINAL SOURCE:
?      ORGANISM: Plasmidium falciparum
?      US-09-210-286-8
```

Query Match	32.0%;	Score 493.5;	DB 4;	Length 921;
Best Local Similarity	40.0%;	Pred. No. 1.8e-40;		
Matches 124; Conservative	39;	Mismatches 94;	Indels 53;	Gaps 15;

Qy	1	DCHPEKSNNGPDMQCCGINIL-VEDPRVCAPPRKQICVHEFLANDNEIKKIOSVN-LKE	58
Db	414	ECYKRET--YSEWTCDESKIKMGQHCACIPPRQKICLHYL-----EKIMNTNENLKY	464
Qy	59	AFISAAAEETFFSWY-YKSKDQSGNGLDKEKLGKGIIPPAFLSMETFGDYRDLFGTD	117
Db	465	AFICAAAEETFLMQNKKKDKNGAELDDELKLGIIIPEDFKQMFTEFYADRYDLCIGTD	524
Qy	118	I-----SKGHGEGSKLKEQIDSLFKNGQDKSPNG-KTROEMWHSHEIWEAMCAL--	168
Db	525	ISSKMDTSKGVG---KVKCNIDVYF-----YKISINSIYRKSMWETNGPVLMEGMLCALSY	577
Qy	169	-----YKIGAKKDDFTFENIGYNNYK-FSKSTTLEEFARPOFLRLTETWYDYCYTRQ	221
Db	578	DTSLNNVPETHAKKLTGEGNNFEFVITFGSOSTTSLKSFSEBQFLRLTLTMEGFECKEOK	637
Qy	222	KYLKDVCDEKCSNDOLACDTECENK---CEDYVK-----YMKKEKKEWIPQD-	265
Db	638	KEYKVLTLAKCKDCD-VQDQDGCNKCKYACAKRQCKQHYHSMIGIWDNKKQKGRTEYVKLI	696
Qy	266	-YYKDERDK 274	
Db	697	PLYEKDKVK 706	

RESULT 8
 US-08-487-826B-16
 Sequence 16, Application US/08487826B
 Patent No. 5993827
 GENERAL INFORMATION:
 APPLICANT: Sim, Kim L.
 APPLICANT: Chluis, Chetan
 APPLICANT: Miller, Louis H.
 APPLICANT: Peterson, David S.
 APPLICANT: Su, Xin-zhaun
 APPLICANT: Wellens, Thomas E.
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/487,826B
 FILING DATE: 10-SEP-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Israelson, Ned
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NIH121.001CPI

```

? TELECOMMUNICATION INFORMATION
? TELEPHONE: (619) 235-8550
?
? TELEFAX: (619) 233-0176
? INFORMATION FOR SEQ ID NO: 16:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2182 amino acids
?
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
? MOLECULE TYPE: protein
?
? HYPOTHEetical: NO
?
? ANTI-SENSE: NO
?
? FRAGMENT TYPE: N-terminal
US-08-487-8266-16

```

Query Match	24.88;	Score 382.5;	DB 2;	Length 2182;
Best Local Similarity	28.08;	Pred. No. 6.4e-29;		
Matches 115;	Conservative 50;	Mismatches 100;	Indels 145;	Gaps 18;

```

Oy 2 CHPKNSNG---:YD:OCCGINIV-----EDPRVCMPRROKLCVHFA-----NOMEI 47
Db 869 CGLKYGGGCKEKPFPNKKCVTPSGVSNATISKDCOALICVPPRRRLRYGGLSOMASNGDET 928
Oy 48 KKLISOV-----NLKAEFKSAAEFFPSWYRK-----SKDG-----80
Db 929 TEVSSEATSAQSSESEKLTAFATIESALETFFLHKHYEKKKPPATQOAGLGVSLPEP 988
Oy 81 --EGNELDKELKE-GRIPFPFLSMYTFEGDYDELFF--GTDISKGH-----123
Db 989 SPPEPDOTOLOQGVYPPFLFLOMEYTLADYLDILYSGSNSDTSPTGKQTPSSNDNIK 1048
Oy 124 -----BESKLEEDISLFKNQDOKSPNGKT-----ROEWTEHS 157
Db 1049 NIVLEAGSGSTEQEEKKK-QIOAKIK---KILNGATSGVPPYTKNSYKTPQOUMENIA 1103
Oy 158 HEIWEAMICAL-----VKI-----GAKKDETFENTGYNNVFKSDEKS- 193
Db 1104 KDIINNAVICALTYKENDARGTSAKIBGNOKLKAALDEANKNMPLEKYQTYNWKLEDSG 1163
Oy 194 -----TTLSEFAKRPQFLMLKEWIDDYTYTTOQTKLADQOEKCSNDOLK-----238
Db 1164 AKSNDTLOPPLKNFVEIPIPEFFMLHBMGNSCFEEFAKKILAOIKHCEMDEDEKOYSGDG 1223
Oy 239 --CD-----TECNKKCEDEYVYMKKK--EKHLPDQKYKXDER 271
Db 1224 EYCEIPIKSOYINVLQDILSSSCAPRCRLYKTEKKEKTEYERKOOKAEVQOK 1273

```

```

1      RESULT 9
2      US-08-568-459A-4
3      ; Sequence 4, Application US/08568459A
4      ; Patent No. 5849306
5      ;
6      ; GENERAL INFORMATION:
7      ;
8      ; APPLICANT: Sim, Kim L.
9      ; APPLICANT: Chitnis, Chetan
10     ; APPLICANT: Miller, Louis H.
11     ; APPLICANT: Peterson, David S.
12     ; APPLICANT: Su, Xin-zhaun
13     ; APPLICANT: Wellemis, Thomas E.
14     ;
15     ; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
16     ; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
17     ;
18     ; NUMBER OF SEQUENCES: 37
19     ;
20     ; CORRESPONDENCE ADDRESS:
21     ;
22     ; ADDRESSEE: Knobbe Martens Olson & Bear
23     ; STREET: 620 Newport Center Drive 16th Floor
24     ; CITY: Newport Beach
25     ; STATE: California
26     ; COUNTRY: US
27     ;
28     ; ZIP: 92660
29     ;
30     ; COMPUTER READABLE FORM:
31     ; MEDIUM TYPE: Floppy disk
32     ;
33     ; COMPUTER: IBM PC compatible
34     ; OPERATING SYSTEM: PC-DOS/MS-DOS
35     ;

```


SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-4

Query Match 17.8%; Score 274; DB 2; Length 1435;
Best Local Similarity 29.2%; Pred. No. 2.2e-18;

Matches 76; Conservative 43; Mismatches 93; Indels 48; Gaps 11;

QY 14 WCCGNINLVEDPRVCMPPRRQKLCVHFLANDNEIKKIQSOVNLKEAFIKSAAEFFSWY 73
DB 474 WECKNPYILSTKDYCVPRROELC---LGNIDRIYD-KNLMITEHILAIAYESILKR 529
QY 74 YKSKDGEENLDEKELKGIKIPAPLASMETFGDYRDLFGDISKGHSGSKLEQID 133
DB 530 KYKNKD-----DKV-----CKIINKTFADIRDIIGTDYNDL-SNRKLVGKIN 573
QY 134 SLKNGDQSPNGKT-ROEWTEHSHEIWEAMCALVIGAKKDDFTENGYNNVFSK 192
DB 574 TNSKYVHRNKNKNDLFRDEWKKVYKKNVNI-----SWVFKDK 612
QY 193 STLEEFAR-POFLRWLTWYDYCYTRQKYLKDVQCKSKNDOLKC-DTECNKCEYD 250
DB 613 TVCKEDDIENIPQFFRFSMGDDYCODKTKMETILKVECK---EKPCDDNCKSKCNSY 669
QY 251 VKYM-KKKKEMIPDKRYKD 269
DB 670 KEMISKKEEYNNQAKOYOE 689

RESULT 10
US-08-487-826B-4

Sequence 4, Application US/08487826B
Patent No. 5993827

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESS: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-487-826B-4

Query Match 17.8%; Score 274; DB 2; Length 1435;
Best Local Similarity 29.2%; Pred. No. 2.2e-18;

Matches 76; Conservative 43; Mismatches 93; Indels 48; Gaps 11;

QY 14 WCCGNINLVEDPRVCMPPRRQKLCVHFLANDNEIKKIQSOVNLKEAFIKSAAEFFSWY 73
DB 474 WECKNPYILSTKDYCVPRROELC---LGNIDRIYD-KNLMITEHILAIAYESILKR 529
QY 74 YKSKDGEENLDEKELKGIKIPAPLASMETFGDYRDLFGDISKGHSGSKLEQID 133
DB 530 KYKNKD-----DKV-----CKIINKTFADIRDIIGTDYNDL-SNRKLVGKIN 573
QY 134 SLKNGDQSPNGKT-ROEWTEHSHEIWEAMCALVIGAKKDDFTENGYNNVFSK 192
DB 574 TNSKYVHRNKNKNDLFRDEWKKVYKKNVNI-----SWVFKDK 612
QY 193 STLEEFAR-POFLRWLTWYDYCYTRQKYLKDVQCKSKNDOLKC-DTECNKCEYD 250
DB 613 TVCKEDDIENIPQFFRFSMGDDYCODKTKMETILKVECK---EKPCDDNCKSKCNSY 669
QY 251 VKYM-KKKKEMIPDKRYKD 269
DB 670 KEMISKKEEYNNQAKOYOE 689

RESULT 11
US-09-210-288-4

Sequence 4, Application US/09210288
Patent No. 6392026

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESS: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-4

Query Match 17.8%; Score 274; DB 4; Length 1435;
Best Local Similarity 29.2%; Pred. No. 2,2e-18;
Matches 76; Conservative 43; Mismatches 93; Indels 48; Gaps 11;

QY 14 WCGGINLVEDPRVCPPPRQKICVHFLANDNEIKLQSOVNLKEAFISAAETFFSWY 73
DB 474 WEKNYIISTDKVCPVPRROELC--LGNIDRIYD-KNLMKEHILAIATIESHLR 529
QY 74 YKSKDEGNELDKELKEKIPAPFLRSMPTGEGDYRDLFGDISKGEGSKLEQID 133
DB 530 KTKNKD-----DKEY-----CKTKNTPADRIDIIIGTDYNNDL-SNKKLVGKIN 573
QY 134 SLEKNDQKSPNGKT-ROBMTWESHSEIWEAMLCALVKGAKKDPTENGYNNVAFSDK 192
DB 574 TNSKYHRKKNDKLFREDKMWIKKDVNNVL-----SWFKKDK 612
QY 193 STLEEFAR-POFLMWLEWIDDYCYTRQKYLKDVQEKSKNDQK-C-DTEKNKCEYD 250
DB 613 TYCKEDDININIPQFFRWFSEWGDYCODYTKMETLKECK---EKPCEDDNCKSKNSY 669
QY 251 VKYM-KKKKEMIPQDKYKD 269
670 KEWISKKEEYKQAKQYOE 689

RESULT 12
US-08-568-459A-21
Sequence 21, Application US/08568459A
Patent No. 5849306

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 311 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
US-08-568-459A-21

Query Match 17.7%; Score 273.5; DB 2; Length 311;
Best Local Similarity 27.5%; Pred. No. 3e-15;
Matches 66; Conservative 22; Mismatches 141; Indels 11; Gaps 4;

QY 28 CMPRRQKICVHFLANDNEIKLQSOVNLKEAFISAAETFFSWYVYKSGEGNELDK 87
DB 24 CMPRRQKICLVYIXX 83
QY 88 ELKEGKIPAPFLRSMPTGEGDYRDLFGDISKGEGSKLEQIDSLFKNDQKSPNGK 147
DB 84 XXXXXX--QFLRSMYTFEGDYRIDICLNTDISKQNDVXXXXXXXXXXXXXXXXX 140
QY 148 TROEMWTEHSEIWEAMLCAL--VKIGAKKDPTENGYNNVAFSDKSTLEEFAR 203
DB 141 SROEMWKTNGPETKQGLCLXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 200
QY 204 QFLRWLTWEDYDCYTRQKYLKDVQEK-C-KSNDQKCDT---CENKKECYVYMKKKKE 259
DB 201 QFLRWLTWEDYDCYTRQKYLKDVQEK-C-KSNDQKCDT---CENKKECYVYMKKKKE 260

RESULT 13
US-08-487-826B-33
Sequence 33, Application US/08487826B
Patent No. 5993827

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 311 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
US-08-487-826B-33

Query Match 17.7%; Score 273.5; DB 2; Length 311;
Best Local Similarity 27.5%; Pred. No. 3e-19;
Matches 66; Conservative 22; Mismatches 141; Indels 11; Gaps 4;

DB 28 CMPPRORLCVHFLANDNEIKKLOSOVNLKEAFKSAAEFFSWYYKSKGEGNELDK 87
|||||
DB 24 CMPPRORLCVHFLANDNEIKKLOSOVNLKEAFKSAAEFFSWYYKSKGEGNELDK 83
OY 88 ELKEGKIPPAFLRSMFYFGDYRDLFGTDSKGGESKLEQIDSLFKNGDOKSPNGK 147
DB 84 XXXXXX--QFLRSMFYFGDYRDLFGTDSKGGESKLEQIDSLFKNGDOKSPNGK 140
OY 148 TROEWTEHSEHWEAMCAL----VKIGAKKDDFTENYGVNKKFSDKSTLLEFAPARP 203
DB 141 SROEWKTINGPEIWGMICALXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 200
OY 204 QFLRMVTEMYDYCYTRQKYLKDVQERC-KSNDLKCPT---EKNKCEDYVYKMKRKE 259
DB 201 QFLRMVTEMYDYCYTRQKYLKDVQERC-KSNDLKCPT---EKNKCEDYVYKMKRKE 260

RESULT 14
US-09-210-288-21
Sequence 21, Application US/09210288
Patent No. 6392026

GENERAL INFORMATION:

APPLICANT: Slim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 311 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
US-09-210-288-21

Query Match 17.7%; Score 273.5; DB 4; Length 311;
Best Local Similarity 27.5%; Pred. No. 3e-19;
Matches 66; Conservative 22; Mismatches 141; Indels 11; Gaps 4;

OY 28 CMPPRORLCVHFLANDNEIKKLOSOVNLKEAFKSAAEFFSWYYKSKGEGNELDK 87
|||||
DB 24 CMPPRORLCVHFLANDNEIKKLOSOVNLKEAFKSAAEFFSWYYKSKGEGNELDK 83
OY 88 ELKEGKIPPAFLRSMFYFGDYRDLFGTDSKGGESKLEQIDSLFKNGDOKSPNGK 147
DB 84 XXXXXX--QFLRSMFYFGDYRDLFGTDSKGGESKLEQIDSLFKNGDOKSPNGK 140
OY 148 TROEWTEHSEHWEAMCAL----VKIGAKKDDFTENYGVNKKFSDKSTLLEFAPARP 203
DB 141 SROEWKTINGPEIWGMICALXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 200
OY 204 QFLRMVTEMYDYCYTRQKYLKDVQERC-KSNDLKCPT---EKNKCEDYVYKMKRKE 259
DB 201 QFLRMVTEMYDYCYTRQKYLKDVQERC-KSNDLKCPT---EKNKCEDYVYKMKRKE 260

RESULT 15
US-08-568-459A-6
Sequence 6, Application US/08568459A
Patent No. 5849306

GENERAL INFORMATION:

APPLICANT: Slim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A

FILED DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29, 655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-6

Query Match 16.9%; Score 260; DB 2; Length 749;
Best Local Similarity 28.7%; Pred. No. 2, 2e-17;
Matches 78; Conservative 41; Mismatches 93; Indels 60; Gaps 13;

QY 14 WCGNINLYEDPR-VCMPPRQKLCVH--FLANDNEIKKIQSOVNLKEAFISAAETFE 70
DB 125 WNCYSNNKTYKPEGVCYCPPRQOLCIGYIFLIRGNEEGIKDHIIN-----KAANYEAMH 178
QY 71 SWYYKSKDQEGNELDKELKEGKIPAPFLRSMYFTPGDYRDFLFGTIDISKGHGEGSKLKE 130
DB 179 LKEKYENAGD-----KICNAIIIGS-----YADIGDIYRGGLDWMRDINT--NKLSE 222
QY 131 QIDSLFKNGD--QKSPNGKTRQEWMTESHSEIWEAMLCALYKIGAKKDDFTENYGYNNY 187
DB 223 KFKIKFMGGNSRKQNDNNERNKMKWKEQKNLIWSSM-----Y 260
QY 188 KESDKSTLE--EFAKRFQFLMLTWTYDDYCYTRQKYLKDYQEKCKSNDQLKC--DTEC 243
DB 261 KHLPKGKTCRHNHNFEEKIPOFLRMLEKMGDEFCDEMGTEYKOLEKICENKN--CSEKKC 317
QY 244 NKKCEDYVKYMKKKK-EMIPQDKYKDERDKK 274
DB 318 KNAQSSYEKWKIKERKNENYNDQSK--KFDSDKK 347

Search completed: April 28, 2003, 10:32:11
Job time : 7.78453 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:32:15 ; Search time 6.16575 Seconds

(without alignments)
3586.886 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554

Perfect score: 1543
Sequence: 1 DCHPRKNSNGYPMQCGNIN.....KKEWTPDKYKDEDRKKRF 276Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Optimal number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	624	40.4	9 US-10-153-273-12	Sequence 12, Appl
2	493.5	32.0	9 US-10-153-273-8	Sequence 8, Appl
3	289.5	18.8	10 US-09-924-154-14	Sequence 14, Appl
4	282	18.3	9 US-09-924-154-16	Sequence 16, Appl
5	274	17.8	1435 US-10-153-273-21	Sequence 21, Appl
6	273.5	17.7	311 US-10-153-273-13	Sequence 13, Appl
7	261	16.9	1421 US-09-924-154-13	Sequence 13, Appl
8	260	16.9	749 US-10-153-273-6	Sequence 6, Appl
9	260	16.9	1086 US-09-924-154-15	Sequence 15, Appl
10	223	14.5	1115 US-10-153-273-2	Sequence 2, Appl
11	218.5	14.2	324 US-10-153-273-17	Sequence 17, Appl
12	210	13.6	700 US-10-153-273-10	Sequence 10, Appl
13	209	13.5	1501 US-09-924-154-17	Sequence 17, Appl
14	164	10.6	277 US-10-153-273-15	Sequence 15, Appl
15	142	9.2	282 US-10-153-273-16	Sequence 16, Appl
16	140	9.1	291 US-10-153-273-13	Sequence 13, Appl
17	133.5	8.7	411 US-10-153-273-19	Sequence 19, Appl
18	126	8.2	411 US-10-153-273-20	Sequence 20, Appl
19	123	8.0	362 US-10-153-273-18	Sequence 18, Appl

20	108	7.0	271	9 US-10-153-273-14	Sequence 14, Appl
21	91	5.9	351	10 US-09-864-761-33332	Sequence 33332, A
22	90	5.8	2332	9 US-09-957-641-2	Sequence 2, Appl
23	90	5.8	2332	9 US-10-187-319-2	Sequence 2, Appl
24	90	5.8	2351	9 US-10-132-829-4	Sequence 4, Appl
25	89	5.8	410	10 US-09-815-242-13393	Sequence 13393, A
26	89	5.8	410	10 US-09-815-242-13628	Sequence 13628, A
27	87	5.6	793	10 US-09-900-237-18	Sequence 18, Appl
28	86	5.6	1139	9 US-09-820-843A-15	Sequence 15, Appl
29	85.5	5.5	1805	9 US-09-820-843A-73	Sequence 73, Appl
30	85	5.5	632	9 US-10-108-605-287	Sequence 287, App
31	85	5.5	956	9 US-10-000-256A-153	Sequence 153, App
32	85	5.5	1261	12 US-10-147-266-2	Sequence 2, Appl
33	84.5	5.5	701	9 US-09-991-496-16	Sequence 16, Appl
34	84.5	5.5	701	9 US-09-874-923-16	Sequence 16, Appl
35	84	5.4	199	9 US-09-864-761-34284	Sequence 34284, A
36	84	5.4	711	9 US-09-298-523B-3	Sequence 3, Appl
37	82.5	5.3	286	9 US-10-106-534-4	Sequence 4, Appl
38	82.5	5.3	503	9 US-09-738-626-5485	Sequence 5485, Ap
39	80	5.2	589	9 US-09-298-523B-14	Sequence 14, Appl
40	80	5.2	690	9 US-09-298-523B-61	Sequence 61, Appl
41	80	5.2	691	9 US-09-298-523B-1	Sequence 1, Appl
42	80	5.2	701	9 US-09-298-523B-62	Sequence 62, Appl
43	80	5.2	905	9 US-10-114-893-127	Sequence 127, App
44	80	5.2	915	10 US-09-880-192-359	Sequence 359, Appl
45	80	5.2	929	9 US-09-298-523B-60	Sequence 60, Appl

ALIGNMENTS

RESULT 1
US-10-153-273-12
Sequence 12, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Shim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhan
Weliems, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21 May 2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 12;
US-10-153-273-12

Query Match 40.4%; Score 624; DB 9; Length 2710;
Best Local Similarity 43.8%; Pred. No. 3.2e-46;
Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11;

QY 1 DCHPKNSNGYPMOC-GNINLVEDPRVCMPRRROKLCVHFANDNEIKKLOSOVNLKE 59
DB 1367 ECPNPKES---YPMDCNNIDISHD-GACMPRRROKLCVYINHEGOTENIKTDMLKDA 1422
60 FIKSAAEFFSWYKSK-DGEGNELDKELKGIIPAFLSMFTYFGDYRDLFGTDI 118
1423 FIKSAAEFFSWYKSKNDSEPAKILDR---GLIPSOFLRSMFTYFGDYRDLFGTDI 1478
QY 119 SKHGSGSKLEQIDSLFKNGDOKSPNGKTRQEWTEHSHETMEALCALYKIGAKKD- 177
DB 1479 SKKQNDYAKAKDGIKFKDQSGSPGLSRQEWMTNPELWKGKLCALYKIVDTDNK 1538
QY 178 --FTWYGNNAKESDK-STLLEEFKRPQFLRWLTWYDYCYTQKYLKDVQEKCS- 233
DB 1539 RKIKNDYSYDKVQSONGNPSELEFAKQFLRMMEWEEECQAEKQKKNENIKKDCNFI 1598
QY 234 NQDKCDT---BCKKCEYVYKMK-KKEWT-----PODKYYK 268
DB 1599 NSTQOCNDAKHCNOCRAVYOEYVENKKKEFGQTNFVLKANVOPDPEYK 1650

RESULT 2

US-10-153-273-8
Sequence 8, Application US/10153273
Patent No. US20020169305A1

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbie Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153/273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FMDV1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:

ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 8;
US-10-153-273-8

Query Match 32.0%; Score 493.5; DB 9; Length 921;
Best Local Similarity 40.0%; Pred. No. 2.8e-35;
Matches 124; Conservative 39; Mismatches 94; Indels 53; Gaps 15;

QY 1 DCHPKNSNGYPMOCGNINL-VEEDPRVCMPRRROKLCVHFANDNEIKKLOSOVNLKE 58
DB 414 ECTKRET---YSEWCDSEKIMGQACIPRRROKLCVH-----EKTMTNNEIKY 464
QY 59 AFIKSAAEFFSWY-YKSKDGEENELDKELKGIIPAFLSMFTYFGDYRDLFGTD 117
DB 465 AFIKSAAEFFLWQNYKKDKGNNAEDDEKIKGIIPEDEKROKFTYFADYRDLFGTD 524
QY 118 I-----SKHGSGSKLEQIDSLFKNGDOKSPNG-KTROEWTEHSHETMEALCAL- 168
DB 535 ISSKRDTSKGVG---KVCNIDIVF---YKINSIRYKFSMWEINGVYIEGMCALASY 577
QY 169 -----VIGAKKDDFTNNGYNNKF--SDKSTLLEFAKRPQFLRWLTWYDYCYTRQ 221
DB 578 DTLNANVPETHKKLTBEGNNNEKVIYFGSDSTYLSKFSERQFLRWLTWEGENPKCKOK 637
QY 222 KYLKDVQEKCSNDQKCDTECNKR---CEDYK-----YAKKKKEWIPDK- 265
DB 638 KEYKVLAKCKDCD-VDDGKCGKGVACQKQCGOYHSGIWIWIDNKKKGRTEYKKI 696
QY 266 -YYDERDKK 274
DB 697 PLYKEDKDK 706

RESULT 3

US-09-924-154-14
Sequence 14, Application US/09924154
Patent No. US20020127241A1

GENERAL INFORMATION:

APPLICANT: Natum, David L.
APPLICANT: Sim, Kim L.
TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
FILE REFERENCE: 05213-0465 43170-262105
CURRENT APPLICATION NUMBER: US/09/924,154
CURRENT FILING DATE: 2001-08-07
PRIOR APPLICATION NUMBER: US 60/223,525
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14

LENGTH: 1143

TYPE: PRT

ORGANISM: Mammalian

US-09-924-154-14

Query Match 18.8%; Score 289.5; DB 10; Length 1143;
Best Local Similarity 28.4%; Pred. No. 3.5e-17;
Matches 80; Conservative 45; Mismatches 100; Indels 57; Gaps 13;

QY 2 CHPKNSNGYPMOCGNINLVEDPRVCMPRRROKLCV---HFLANDNEIKKLOSOVNLKE 58
DB 462 CNK--SNNEKSNMNCIGTFNKFPGTCEPPRQTLCLIGRTYLLRHGE-----EDYKE 512

QY 59 AFIKSAAEFFSWYKSGDKGNELEKELKELKIPPAFLNSMFTYFGDYRDLFGT 118
DB 513 HILGASIVAEQOLTKYKREKEN-----ALCSITONSADLADIINGS 557
QY 119 SKGHGSGSLKQIDSLFNGDOKSPNGKT-ROEWTEHSHEIWEAMLCALVIGAKKDD 177
DB 558 IKDY-YGKKMEENLNKVNMDKRNNEESLKIIFREKMWDEKENVKMSAVL-----KNKE 611
QY 178 FTEYGYNNVWFSDKSTLEEFKARPOFLRWLTWYDYCYTROKYL---KDYQKCKSN 234
DB 612 TCKDY-----DKFOKITQFLRWTFEWMGDDDCERKREKIYSPESKVECKK 657
QY 235 DOLKCD-TECNKKCEDYKVMK-KKKEMIPQ-DKYKDERDK 273
DB 658 D--CDENTCKNCKSEYKKWIDLKSEYKQVDKTKDKKK 696

RESULT 4

US-09-924-154-16
Sequence 16, Application US/09924154
Patent No. US20020127241A1

GENERAL INFORMATION:

APPLICANT: Natum, David L.
APPLICANT: Sim, Kim L.
TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
FILE REFERENCE: 05213-0465 43170-262105
CURRENT APPLICATION NUMBER: US/09/924,154
CURRENT FILING DATE: 2001-08-07
PRIOR APPLICATION NUMBER: US 60/223,525
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 16
LENGTH: 972
TYPE: prf
ORGANISM: Mammalian
US-09-924-154-16

Query Match 18.3%; Score 282; DB 10; Length 972;
Best Local Similarity 27.8%; Pred. No. 1,3e-16;

Matches 84; Conservative 45; Mismatches 105; Indels 68; Gaps 12;

QY 2 CHPKNSNGYPMQGNINLVEDPRVCMPPRRKOLCVHFLAND-----NEIKKLSQVNL 56
DB 138 CEKGNDR--TWQCINEHIKDFDVCPPRRQOLCLGNLDKDEFKVNLDKFLNRI-- 192
QY 57 KEAFIKSAAEFFSWYKSGDKGNELEKELKELKIPPAFLNSMFTYFGDYRDLFGT 116
DB 193 ----ILGIDEGKFLLEKTKRNKHNEMNYLDERA-----CKYLNYSDDYKNTILK 239
QY 117 D-----ISKGHSGSLKQIDSLFNGDOKSPNGKTROEWTEHSHEIWEAM 164
DB 240 DMRDONSITENILKGNFEG--IKANYISMYSADLSLD-EFRKHWMDONKQOLWEAL 236
QY 165 LCAALVIGAKKDDFTENYGNVYKFSKSTLEEFKARPOFLRWLTWYDYCYTROKYL 224
DB 297 SC-----EFGKGNHT-----GYCLMEDNDNQYLHMPFEKMKDPCIDILKMN 338
QY 225 KDVQKCK-----KSNQDLKCDTECNKKCEDYKVMK-KKKEMIPQDKYKDERDK 274
DB 339 DYKKEPCIDIKYKSPRSENPSPDVATVCKNSCTDYDKMIINRKREYKMSOSKYK--RDRS 396
QY 275 RF 276
DB 397 LF 398

RESULT 5
US-10-153-273-4
Sequence 4, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhuan
Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSER: Knobe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH21.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-153-273-4
Query Match 17.8%; Score 274; DB 9; Length 1435;
Best Local Similarity 29.2%; Pred. No. 1.1e-15;
Matches 76; Conservative 43; Mismatches 93; Indels 48; Gaps 11;
QY 14 WQGNINLVDPVCMPPRRKOLCVHFLANDNEIKKLSQVNLKEAFIKSAAEFFSWY 73
DB 474 WECKNPYILSTKDYCVPRQOELC--LGNIDRIYD-KNLMKKEHLAIAIESRILK 539
QY 74 YKSKDGEENLELKEKGIIPPAFLNSMFTYFGDYRDLFGTDISKGHSGSLKQID 133
DB 530 KYRKNK-----DKV-----CKIINKTRADIDITIGGTDYNDL-SNRKLKGIN 573
QY 134 SLEKNGDQKSPNGKT-ROEWTEHSHEIWEAMLCALVIGAKKDDFTENYGNVYKFSK 192
DB 574 TNSKYVHRNKNKDLFDEWKMVIKKDVMVYI-----SWYKDK 612
QY 193 STLEEFKAR-POFLRWLTWYDYCYTROKYLKDVQKCKSDQKLC-DTECNKKCEDY 230
DB 613 TVCKEDDIENIPQFFRFWSMGDDYCODKTKMTETLVECK---EKPCEDNCKSKNSY 669
QY 251 VKYM-KKKEMIPQDKYK 269
DB 670 KEMISKKEEYNNQAKOYOE 689

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH21.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEetical: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-153-273-6

Query Match 16.9%; Score 260; DB 9; Length 749;
Best Local Similarity 28.7%; Pred. No. 8.5e-15;
Matches 78; Conservative 41; Mismatches 93; Indels 60; Gaps 13;

QY 14 WCGNINLVDP--VCMPPRRKLCVH--FLANDNEIKIQSOVNLKFAFKSAATFF 70
DB 125 WNCYSNNKVKRPGVCGPPRQOLCGYIFLRDGNBEGIKDHN-----KAANYEAMH 178
QY 71 SWYYSKQEGNDELKELKEGKIPAFILSMFTYFGDYDFLFGTIDISKHGBGSLKE 130
DB 179 LKEEYENAGD-----KICNAILGS---YADIGDIYAGLDVWDINT--NKLSE 222
QY 131 QIDSLFRNGD---OKSPNGKTRQEWTEHSEIWEAMLCALVKGAKKDDFTENYGYNNV 187
DB 223 KFKITWGGNSRKQNDNERNKRWKRNLIWSS-----V 260
QY 188 KFSDKSTTLE--EFAKRPQFLMLTEWYDYCYTRKYLKDVQCKCKSNDQKLC-DTEC 243
DB 261 KHIPKGTCTCRHNNFEKIPQFLRWLKEWGEFCEMGTEVKOLEKICENKN---CSEKCK 317
244 NKCEDEVKMKKK-EMIPQDKYKDERDK 274
DB 318 KNACSSYEKWKERKNEYNLQSK--KFSDDK 347

RESULT 9
US-09-924-154-15
Sequence 15, Application US/09924154
Patent No. US20020127241A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
FILE REFERENCE: 05213-0465 43170-262105
CURRENT APPLICATION NUMBER: US/09/924,154
PRIOR FILING DATE: 2001-08-07
PRIOR APPLICATION NUMBER: US 60/223,525
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin version 3.1
SEQ ID NO 15
LENGTH: 1086
TYPE: PRT
ORGANISM: Mammalian
US-09-924-154-15

Query Match 16.9%; Score 260; DB 10; Length 1086;
Best Local Similarity 28.7%; Pred. No. 1.3e-14;
Matches 78; Conservative 41; Mismatches 93; Indels 60; Gaps 13;

QY 14 WCGNINLVDP--VCMPPRRKLCVH--FLANDNEIKIQSOVNLKFAFKSAATFF 70
DB 367 WNCYSNNKVKRPGVCGPPRQOLCGYIFLRDGNBEGIKDHN-----KAANYEAMH 420
QY 71 SWYYSKQEGNDELKELKEGKIPAFILSMFTYFGDYDFLFGTIDISKHGBGSLKE 130
DB 421 LKEEYENAGD-----KICNAILGS---YADIGDIYAGLDVWDINT--NKLSE 464
QY 131 QIDSLFRNGD---OKSPNGKTRQEWTEHSEIWEAMLCALVKGAKKDDFTENYGYNNV 187
DB 465 KFKITWGGNSRKQNDNERNKRWKRNLIWSS-----V 502
QY 188 KFSDKSTTLE--EFAKRPQFLMLTEWYDYCYTRKYLKDVQCKCKSNDQKLC-DTEC 243
DB 503 KHIPKGTCTCRHNNFEKIPQFLRWLKEWGEFCEMGTEVKOLEKICENKN---CSEKCK 559
QY 244 NKCEDEVKMKKK-EMIPQDKYKDERDK 274
DB 560 KNACSSYEKWKERKNEYNLQSK--KFSDDK 589

RESULT 10
US-10-153-273-2
Sequence 2, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chilens, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSER: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH21.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1115 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium vivax
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-153-273-2

Query Match 14.5%; Score 223; DB 9; Length 1115;
Best Local Similarity 26.1%; Pred. No. 2.5e-11;

Matches 70; Conservative 44; Mismatches 90; Indels 64; Gaps 14;

13 DMOCSNINLVEDPRVCMPRRKCKLVHFLAN-----DNEIKRLOSQVNLKEAFI 61
Db 272 DWDG---NFKRD--VCIDPRYQJCMKELINLVNNTDTHFHDITFRKLY---LKRRLI 322
QY 62 KSAAEFTFSYVYKSKDSEGNELDKELKEGIPPAFLRSMEFTYFGDRFLFGTDSIGK 121
Db 323 YDAVE-----GDLTKLNRYNKDFCKDIRMSLGPGLIMGTDA-EG 366
QY 122 HGECSKLEQIDSLFKNGDOKSPNGKTRQEWTHSHETHEAMICALVKGAKKDPFTEN 181
367 IGYSKVVENNRSTF-GTDEKAQ--QRRKQWNSKAQIWTAMMYSVK--RLKGNFTWI 421
QY 182 YGVN---NVKESDASTLEEFARPOFLRLMTEWYDDCYTRQKYLKDQVEKGS----N 234
Db 422 CKLVAVNIE-----POLYRMIREGRDYSELPLEYQKLEKCDGKINYP 467
QY 235 DOLKCDI-ECKNCKEDYVKT-KKKKEW 260
Db 468 DKVKCYVPCOMACKSYDOMITRRKNOW 495

RESULT 11
US-10-153-273-17
Sequence 17, Application US/10153273
Patent No. US20020169305A1

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun

Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

NUMBER OF SEQUENCES: 37 AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear

STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/153,273

FILING DATE: 21-May-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/210,288

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Fuller, Michael

REGISTRATION NUMBER: 36,516

REFERENCE/DOCKET NUMBER: NIH21.1FMDV1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 324 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:

SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-153-273-17

Query Match 14.2%; Score 218.5; DB 9; Length 324;
Best Local Similarity 24.7%; Pred. No. 1.4e-11;

Matches 60; Conservative 15; Mismatches 155; Indels 13; Gaps 2;

28 CMPPRRKCKLVHFLANDNEIKRLOSQVNLKEAFISAAEFTFSYVYKSKDSEGNELDK 87
Db 25 CIPPRKCKLVHFLAXX 84
QY 88 ELKEGKIPPAFLRSMEFTYFGDRFLFGTDSIGKSGEGSKLEQIDSLFKNGDOKSPNGK 147
Db 85 XXXXXD---FRRQMEFTFADYRDICLGTDSKKDTSXXXXXXXXXXXXXXXXXINSIR 140
QY 148 TROEWTHSHETHEAMICAL-----VKIGAKKDPFTENGYNNVKSFSKSTLEE 198
Db 141 YKSMWETNGPYWGMICALXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 200
QY 199 FAKRPOFLRLMTEWYDDCYTRQKYLKDQVEKCSNDQLKCDTECNKCKEDYVYMKKK 258
Db 201 XXXRPOFLRLMTEWENCKDQKKEYVLLAKCXXXXXXXXXXCVACAKDQCKOYH 260
QY 259 EMI 261
Db 261 SWI 263

RESULT 12
US-10-153-273-10

Sequence 10, Application US/10153273
Patent No. US20020169305A1

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun

Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

NUMBER OF SEQUENCES: 37 AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear

STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/153,273

FILING DATE: 21-May-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/210,288

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Fuller, Michael

REGISTRATION NUMBER: 36,516

	Query Match	13.5%	Score 209;	DB 10;	Length 1501;	
	Best Local Similarity	22.6%;	Pred. No.	6,1e-10;		
	Matches 65;	Conservative 55;	Mismatches 115;	Indels 52;	Gaps 13;	
Oy	2 CHPKKNSNGYPMOC--GNINLYVEDPVMCPRRQKLC---VHFLTANDNEIKKLDSQVNL 56			:	::: :	: :

```

1      RESULT 14
2      US-10-153-273-15
3      : Sequence 15, Application US/10153273
4      : Patent No. US20020169305A1
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: Sim, Kim L.
9      :           Chitnis, Chetan
10     :           Miller, Louis H.
11     :           Peterson, David S.
12     :           Su, Xin-zhaun
13     :           Wellens, Thomas E.
14     :
15     : TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
16     : AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
17     :
18     : NUMBER OF SEQUENCES: 37
19     :
20     : CORRESPONDENCE ADDRESS:
21     : ADDRESSEE: Knobbe Martens Olson & Bear
22     : STREET: 620 Newport Center Drive 16th Floor
23     : CITY: Newport Beach
24     : STATE: California
25     : COUNTRY: US
26     : ZIP: 92660
27     :
28     : COMPUTER READABLE FORM:
29     : MEDIUM TYPE: Floppy disk
30     : COMPUTER: IBM PC compatible
31     : OPERATING SYSTEM: PC-DOS/MS-DOS
32     : SOFTWARE: Patentin Release #1.0, Version #1.25.
33     :
34     : CURRENT APPLICATION DATA:
35     : APPLICATION NUMBER: US/10/153,273
36     : FILING DATE: 21-May-2002
37     : CLASSIFICATION: <Unknown>
38     :
39     : PRIOR APPLICATION DATA:
40     : APPLICATION NUMBER: US/09/210,288
41     : FILING DATE: <Unknown>
42     : ATTORNEY/AGENT INFORMATION:
43     : NAME: Fuller, Michael
44     :
45     : REGISTRATION NUMBER: 36,516
46     : REFERENCE/DOCKET NUMBER: NIH121.1fwdv1
47     :
48     : TELECOMMUNICATION INFORMATION:
49     : TELEPHONE: (619) 235-8550
50     : TELEFAX: (619) 235-0176
51     :
52     : INFORMATION FOR SEQ ID NO: 15:
53     :
54     : SEQUENCE CHARACTERISTICS:
55     : LENGTH: 277 amino acids
56     : TYPE: amino acid
57     : STRANDEDNESS: single
58     : TOPOLOGY: linear
59     :
60     : MOLECULE TYPE: peptide
61     :
62     : HYPOTHETICAL: NO
63     :
64     : ANTI-SENSE: NO
65     :
66     : FRAGMENT TYPE: Internal
67     :
68     : ORIGINAL SOURCE:
69     :
70     : SEQUENCE DESCRIPTION: SEQ ID NO: 15:
71     :
72     :
73     :
74     :
75     :
76     :
77     :
78     :
79     :
80     :
81     :
82     :
83     :
84     :
85     :
86     :
87     :
88     :
89     :
90     :
91     :
92     :
93     :
94     :
95     :
96     :
97     :
98     :
99     :
100    :
101    :
102    :
103    :
104    :
105    :
106    :
107    :
108    :
109    :
110    :
111    :
112    :
113    :
114    :
115    :
116    :
117    :
118    :
119    :
120    :
121    :
122    :
123    :
124    :
125    :
126    :
127    :
128    :
129    :
130    :
131    :
132    :
133    :
134    :
135    :
136    :
137    :
138    :
139    :
140    :
141    :
142    :
143    :
144    :
145    :
146    :
147    :
148    :
149    :
150    :
151    :
152    :
153    :
154    :
155    :
156    :
157    :
158    :
159    :
160    :
161    :
162    :
163    :
164    :
165    :
166    :
167    :
168    :
169    :
170    :
171    :
172    :
173    :
174    :
175    :
176    :
177    :
178    :
179    :
180    :
181    :
182    :
183    :
184    :
185    :
186    :
187    :
188    :
189    :
190    :
191    :
192    :
193    :
194    :
195    :
196    :
197    :
198    :
199    :
200    :
201    :
202    :
203    :
204    :
205    :
206    :
207    :
208    :
209    :
210    :
211    :
212    :
213    :
214    :
215    :
216    :
217    :
218    :
219    :
220    :
221    :
222    :
223    :
224    :
225    :
226    :
227    :
228    :
229    :
230    :
231    :
232    :
233    :
234    :
235    :
236    :
237    :
238    :
239    :
240    :
241    :
242    :
243    :
244    :
245    :
246    :
247    :
248    :
249    :
250    :
251    :
252    :
253    :
254    :
255    :
256    :
257    :
258    :
259    :
260    :
261    :
262    :
263    :
264    :
265    :
266    :
267    :
268    :
269    :
270    :
271    :
272    :
273    :
274    :
275    :
276    :
277    :
278    :
279    :
280    :
281    :
282    :
283    :
284    :
285    :
286    :
287    :
288    :
289    :
290    :
291    :
292    :
293    :
294    :
295    :
296    :
297    :
298    :
299    :
300    :
301    :
302    :
303    :
304    :
305    :
306    :
307    :
308    :
309    :
310    :
311    :
312    :
313    :
314    :
315    :
316    :
317    :
318    :
319    :
320    :
321    :
322    :
323    :
324    :
325    :
326    :
327    :
328    :
329    :
330    :
331    :
332    :
333    :
334    :
335    :
336    :
337    :
338    :
339    :
340    :
341    :
342    :
343    :
344    :
345    :
346    :
347    :
348    :
349    :
350    :
351    :
352    :
353    :
354    :
355    :
356    :
357    :
358    :
359    :
360    :
361    :
362    :
363    :
364    :
365    :
366    :
367    :
368    :
369    :
370    :
371    :
372    :
373    :
374    :
375    :
376    :
377    :
378    :
379    :
380    :
381    :
382    :
383    :
384    :
385    :
386    :
387    :
388    :
389    :
390    :
391    :
392    :
393    :
394    :
395    :
396    :
397    :
398    :
399    :
400    :
401    :
402    :
403    :
404    :
405    :
406    :
407    :
408    :
409    :
410    :
411    :
412    :
413    :
414    :
415    :
416    :
417    :
418    :
419    :
420    :
421    :
422    :
423    :
424    :
425    :
426    :
427    :
428    :
429    :
430    :
431    :
432    :
433    :
434    :
435    :
436    :
437    :
438    :
439    :
440    :
441    :
442    :
443    :
444    :
445    :
446    :
447    :
448    :
449    :
450    :
451    :
452    :
453    :
454    :
455    :
456    :
457    :
458    :
459    :
460    :
461    :
462    :
463    :
464    :
465    :
466    :
467    :
468    :
469    :
470    :
471    :
472    :
473    :
474    :
475    :
476    :
477    :
478    :
479    :
480    :
481    :
482    :
483    :
484    :
485    :
486    :
487    :
488    :
489    :
490    :
491    :
492    :
493    :
494    :
495    :
496    :
497    :
498    :
499    :
500    :
501    :
502    :
503    :
504    :
505    :
506    :
507    :
508    :
509    :
510    :
511    :
512    :
513    :
514    :
515    :
516    :
517    :
518    :
519    :
520    :
521    :
522    :
523    :
524    :
525    :
526    :
527    :
528    :
529    :
530    :
531    :
532    :
533    :
534    :
535    :
536    :
537    :
538    :
539    :
540    :
541    :
542    :
543    :
544    :
545    :
546    :
547    :
548    :
549    :
550    :
551    :
552    :
553    :
554    :
555    :
556    :
557    :
558    :
559    :
560    :
561    :
562    :
563    :
564    :
565    :
566    :
567    :
568    :
569    :
570    :
571    :
572    :
573    :
574    :
575    :
576    :
577    :
578    :
579    :
580    :
581    :
582    :
583    :
584    :
585    :
586    :
587    :
588    :
589    :
590    :
591    :
592    :
593    :
594    :
595    :
596    :
597    :
598    :
599    :
600    :
601    :
602    :
603    :
604    :
605    :
606    :
607    :
608    :
609    :
610    :
611    :
612    :
613    :
614    :
615    :
616    :
617    :
618    :
619    :
620    :
621    :
62
```


Result	No.	Score	Query Match	Length	DB	ID	Description
	1	276	100.0	3542	22	AA662142	P. falciparum FCRI
	2	17	6.2	407	22	AA662148	P. falciparum varc
	3	12	4.3	351	22	AA662149	P. falciparum varc
	4	11	4.0	308	22	AA662151	P. falciparum varc
	5	11	4.0	311	22	AA662150	P. falciparum varc
	6	11	4.0	324	21	AA777911	Plasmodium DBL ger
	7	11	4.0	793	16	AA770234	P. falciparum E31a
	8	11	4.0	921	18	AAW22480	Plasmodium E31a
	9	11	4.0	921	21	AA777902	P. falciparum eb1
	10	10	3.6	32	22	AA662143	P. falciparum FCRI

DD 1339 INSAWAEI FSWI I INSDGEGNELDKRELEAGK I PPAF LKSMF YTFGDIRDF LRGTDISK 1339

QY 121 GHGEGSKLEQIDSLFKNGDQKSPNGKTRQEWTTESHSHEIWEAMLCALYKIGAKKDDFTE 180

Db 1399 GHEGSSKEKEQIDSLFKNGDOKSPNGKTRQEWTEHSEIWFAMICALVKGAKKDFTE 1456
 QY 181 NGVNNVNFSDKSTLEEFAPKPOFLRWLTETMYDYCYTRQKYLKDVQEKKSNDQKCD 240
 Db 1459 NGVNNVNFSDKSTLEEFAPKPOFLRWLTETMYDYCYTRQKYLKDVQEKKSNDQKCD 1518
 QY 241 TECKNKCEDYVYMKKKKEMIPQDKYKXDERDKRNF 276
 Db 1519 TECKNKCEDYVYMKKKKEMIPQDKYKXDERDKRNF 1554

RESULT 2

AAB62148
 ID AAB62148 standard; peptide: 407 AA.
 AC AAB62148;
 XX
 XX
 DT 29-MAY-2001 (first entry)
 XX

P. falciparum varCSA polypeptide A4 DBL4-gamma.

KW FCR3.varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
 KM erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
 XX malaria; protozoacide; A4 DBL4-gamma.

OS Plasmodium falciparum.

WO200116326-A2.

08-MAR-2001.

01-SEP-2000; 2000WO-US24195.

01-SEP-1999; 99US-0152023.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
 PI Gysin J, Pouvelle B, Fujil N, Smith J;
 DR WPI; 2001-235109/24.

XX Novel FCR3.varCSA protein, useful for modulating parasitized red blood

PT cell binding, sequestration and onset of maternal malaria -
 PS Disclosure; Page 72-73; 78pp; English.

XX The invention relates to a P. falciparum FCR3.varCSA protein, that is

CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
 CC corresponding P.falciparum erythrocyte membrane protein 1 (PFEMP1)
 CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The

CC protein and the encoding gene are useful for treating and preventing
 CC maternal malaria in a patient afflicted with maternal malaria. The
 CC present sequence represents a P. falciparum varCSA polypeptide
 CC A4 DBL4-gamma.

SQ Sequence 407 AA;

Query Match 6.2%; Score 17; DB 22; Length 407;
 Best Local Similarity 100.0%; Pred. No. 1.6e-08;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 GYRDFLEFCTDISKGGH 123
 Db 128 GYRDFLEFCTDISKGGH 144

RESULT 3

AAB62149
 ID AAB62149 standard; peptide: 351 AA.
 XX

AC AAB62149;

DT 29-MAY-2001 (first entry)

P. falciparum varCSA polypeptide A4resDBL3-gamma.

KW FCR3.varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
 KM erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
 XX malaria; protozoacide; A4resDBL3-gamma.

OS Plasmodium falciparum.

WO200116326-A2.

08-MAR-2001.

01-SEP-2000; 2000WO-US24195.

01-SEP-1999; 99US-0152023.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
 PI Gysin J, Pouvelle B, Fujil N, Smith J;
 DR WPI; 2001-235109/24.

PT Novel FCR3.varCSA protein, useful for modulating parasitized red blood
 XX cell binding, sequestration and onset of maternal malaria -
 PS Claim 54; Page 73-74; 78pp; English.

CC The invention relates to a P. falciparum FCR3.varCSA protein, that is
 CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
 CC corresponding P.falciparum erythrocyte membrane protein 1 (PFEMP1)
 CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
 CC protein and the encoding gene are useful for treating and preventing
 CC maternal malaria in a patient afflicted with maternal malaria. The
 CC present sequence represents a P. falciparum varCSA polypeptide
 CC A4resDBL3-gamma.

SQ Sequence 351 AA;

Query Match 4.3%; Score 12; DB 22; Length 351;
 Best Local Similarity 100.0%; Pred. No. 0.0015;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 EAFIKSAAAEF 69
 Db 70 EAFIKSAAAEF 81

RESULT 4

AAB62151
 ID AAB62151 standard; peptide: 308 AA.

AC AAB62151;

DT 29-MAY-2001 (first entry)

P. falciparum varCSA polypeptide ITG2-CS2 DBL2.

KW FCR3.varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
 KM erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
 XX malaria; protozoacide; ITG2-CS2 DBL2..

OS Plasmodium falciparum.

WO200116326-A2.

08-MAR-2001.

PF 01-SEP-2000; 2000MO-US24195.
XX 01-SEP-1999; 99US-0152023.
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
PI Gysin J, Pouvelle B, Fujii N, Smith J;
XX WPI; 2001-235109/24.
DR
XX Novel FCRA varCSA protein, useful for modulating parasitized red blood
PT cell binding, sequestration and onset of maternal malaria -
XX
PS Claim 54; Page 75-76; 78pp; English.
XX
XX The invention relates to a P. falciparum FCRA varCSA protein, that is
CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
CC corresponding P.falciparum erythrocyte membrane protein 1 (PFEMP1)
modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
protein and the encoding gene are useful for treating and preventing
maternal malaria in a patient afflicted at a risk for contracting
maternal malaria or in a patient afflicted with maternal malaria. The
present sequence represents a P. falciparum varCSA polypeptide
IG2-CS2 DBL2.

XX
SQ Sequence 308 AA;

Query Match 4.0%; Score 11; DB 22; Length 308;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 CMPPRORLKV 38
|||
Db 34 CMPPRORLKV 44

RESULT 5
AAB62150

ID AAB62150 standard; peptide; 311 AA.

XX
AC AAB62150;

XX
DT 29-MAY-2001 (first entry)

XX
DE P. falciparum varCSA polypeptide FCRA var3DBL-gamma.

XX
XX FCRA varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
malaria; protozoacide; FCRA var3DBL-gamma.

XX
XX Plasmodium falciparum.

XX
OS WO200116326-A2.

XX
PN 08-MAR-2001.

XX
PD 01-SEP-2000; 2000MO-US24195.

XX
PE 01-SEP-1999; 99US-0152023.

XX
PR (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX
PA Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
PI Gysin J, Pouvelle B, Fujii N, Smith J;
XX WPI; 2001-235109/24.

XX
DR Novel FCRA varCSA protein, useful for modulating parasitized red blood

XX
PT cell binding, sequestration and onset of maternal malaria -
XX
PS Disclosure; Page 74-75; 78pp; English.

CC The invention relates to a P. falciparum FCRA varCSA protein, that is
CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
CC corresponding P.falciparum erythrocyte membrane protein 1 (PFEMP1)
modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
protein and the encoding gene are useful for treating and preventing
maternal malaria in a patient afflicted at a risk for contracting
maternal malaria or in a patient afflicted with maternal malaria. The
present sequence represents a P. falciparum varCSA polypeptide
FCRA var3DBL-gamma.

XX
SQ Sequence 311 AA;

Query Match 4.0%; Score 11; DB 22; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 CMPPRORLKV 38
|||
Db 36 CMPPRORLKV 46

RESULT 6

ID AAY7911 standard; Protein; 324 AA.

XX
AC AAY7911;

XX
DT 13-JUN-2000 (first entry)

XX
DE Plasmodium DBL gene family E3a conserved domain.

XX
KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
DABP; Sialic Acid Binding Protein; SAMP; malaria; vaccine; immunisation;
protozoacide; E3a.

XX
XX Plasmodium sp.

XX
OS Plasmodium sp.

XX
XX Key Location/Qualifiers

XX
FT Misc-difference 1..324

XX
FT /note= "residues indicated Xaa are unspecified"

XX
PN US593827-A.

XX
PD 30-NOV-1999.

XX
PE 07-JUN-1995; 95US-0487826.

XX
PF 10-SEP-1993; 93US-0119677.

XX
PR (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX
PA Slim KL, Chlens C, Peterson DS, Su X, Wellens TE, Miller LH;

XX
PI WPI; 2000-194198/17.

XX
PT Isolated protein binding domains from Plasmodium vivax and Plasmodium

XX
PT malaria -
PT erythrocyte binding proteins useful for vaccinating against

XX
PT malaria -
PT Examples; Fig 1; 93pp; English.

XX
PS The invention relates to ebl-1 polypeptides that are encoded by the DBL

XX
XX (Duffy-binding like) gene family. The ebl-1 proteins are substantially

XX
XX identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid

XX
XX Binding Protein (SAMP), which are soluble proteins that appear in the

XX
XX culture supernatant after erythrocytes infected with malaria release

XX
XX macrophages. Immunological studies indicate that DABP and SAMP are the

XX
XX respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy

XX
XX and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be

XX
XX used to vaccinate against malaria, especially caused by P. falciparum.

XX
XX Immunization with the polypeptide provides effective protection against

XX
XX malaria.


```

DE P. falciparum ebl-1 related polypeptide.
XX DBI gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
XX DABP; Stalic Acid Binding Protein; SABB; malaria; vaccine; Immunisation;
XX protozoacide.
XX
XX OS Plasmodium falciparum.
XX
XX US593827-A.
XX
XX 30-NOV-1999.
XX
XX 07-JUN-1995; 95US-0487826.
XX
XX 10-SEP-1993; 93US-0119677.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Slim XL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;
XX
XX WPI; 2000-194198/17.
XX
XX N-PSDB; AA298284.
XX
XX Isolated protein binding domains from Plasmodium vivax and Plasmodium
XX falciparum erythrocyte binding proteins useful for vaccinating against
XX malaria -
XX
XX Disclosure; Columns 61-66; 93bp; English.
XX
XX
XX The invention relates to ebl-1 polypeptides that are encoded by the DBI
XX (Duffy-binding like) gene family. The ebl-1 proteins are substantially
XX identical to the Duffy Antigen Binding Protein (DABP) and Stalic Acid
XX Binding Protein (SABB), which are soluble proteins that appear in the
XX culture supernatant after erythrocytes infected with malaria release
XX merozoites. Immunochemical studies indicate that DABP and SABB are the
XX respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
XX and stalic acid receptors on erythrocytes. The ebl-1 polypeptides may be
XX used to vaccinate against malaria, especially caused by P. falciparum.
XX Immunization with the polypeptide provides effective protection against
XX malaria.
XX
XX
XX Sequence 921 AA;
XX
XX
XX Query Match 4.0%; Score 11; DB 21; Length 921;
XX Best Local Similarity 100.0%; Pred. No. 0.036;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 202 RPOFLRWLLEW 212
XX |||||||||
XX 618 RPOFLRWLLEW 628
XX
XX
XX RESULT 10
XX AAB62143
XX AAB62143 standard; peptide: 32 AA.
XX
XX AAB62143;
XX
XX 29-MAY-2001 (first entry)
XX
XX P. falciparum FCR3.varCSA protein fragment.
XX
XX FCR3.varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
XX erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
XX malaria; protozoacide.
XX
XX Plasmodium falciparum.
XX
XX OS
XX WO200116326-A2.
XX
XX
XX 08-MAR-2001.
XX
XX
XX 01-SEP-2000; 2000WO-US24195.
XX
XX

```

XX	01-SEP-1999;	99US-0152023.
PR	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
XX	Scherf A, Miller LH, Gamain B, Baruch DJ, Buffet P, Scheldig C;	
PI	Gysin J, Pouvelle B, Fujii N, Smith J;	
XX	WPI; 2001-235109/24.	
DR		
XX	Novel FCRA3.varcSA protein, useful for modulating parasitized red blood	
PT	cell binding, sequestration and onset of maternal malaria -	
PT	Claim 25; Page 21; 78pp: English.	
XX		
XX	The invention relates to a P. falciparum FCRA3.varcSA protein, that is	
CC	capable of binding to chondroitin sulfate A (CSA). The var gene and the	
CC	corresponding P. falciparum erythrocyte membrane protein 1 (PFEMP1)	
CC	modulate adhesion of parasitized red blood cell (PRBC) to CSA. The	
CC	protein and the encoding gene are useful for treating and preventing	
CC	maternal malaria in a patient identified at a risk for contracting	
CC	maternal malaria or in a patient afflicted with maternal malaria. The	
CC	present sequence represents a fragment of the P. falciparum FCRA3.varcSA	
CC	protein.	
XX		
SO	Sequence 32 AA;	
Query Match	3.6%; Score 10; DB 22; Length 32;	
Best Local Similarity	100.0%; Pred. No. 0.019;	
Matches 10; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
OY	102 MEYTFG DYRD 111	
Db	19 MEYTFG DYRD 28	
RESULT 11		
AA777915		
ID	AA777915 standard; Protein; 311 AA.	
XX		
AC	AA777915;	
XX		
DT	13-JUN-2000 (first entry)	
XX		
DE	Plasmodium DBL gene family Proj3 conserved domain F3.	
XX		
KW	DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;	
KW	DABP; Sialic Acid Binding Protein; SAMP; malaria; vaccine; Immunisation;	
KM	protozoacide; Proj3.	
XX		
OS	Plasmodium sp.	
XX		
FT	Key location/Qualifiers	
FT	Misc-difference 1..311	
XX	/note- "residues indicated Xaa are unspecified"	
XX		
PN	US5933827-A.	
XX		
PD	30-NOV-1999.	
XX		
PF	07-JUN-1995; 95US-0487826.	
XX		
PR	10-SEP-1993; 93US-0119677.	
XX		
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
XX		
PI	Sim KL, Chitnis C, Peterson DS, Su X, Wellem TE, Miller LH;	
XX	WPI; 2000-194198/17.	
DR		
XX		
FT	Isolated protein binding domains from Plasmodium vivax and Plasmodium	
FT	falciparum erythrocyte binding proteins useful for vaccinating against	
PT	malaria -	

XX PS Examples; Fig 1; 93pp; English.
 CC CC The invention relates to ebl-1 polypeptides that are encoded by the DBL
 CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
 CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
 CC Binding Protein (SABP), which are soluble proteins that appear in the
 CC culture supernatant after erythrocytes infected with malaria release
 CC merozoites. Immunochemical studies indicate that DABP and SABP are the
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
 CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
 CC used to vaccinate against malaria, especially caused by P. falciparum.
 CC Immunization with the polypeptide provides effective protection against
 CC malaria.
 CC XX
 SQ Sequence 311 AA;
 Query Match 3.6%; Score 10; DB 21; Length 311;
 Best Local Similarity 100.0%; Pred. No. 0.14;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 28 CMPPRRQKLC 37
 |||||
 Db 24 CMPPRRQKLC 33
 RESULT 12
 AAR70236
 ID AAR70236 standard; Protein; 2703 AA.
 XX AC
 XX AAR70236;
 XX DT 22-SEP-1995 (first entry)
 XX DE P. falciparum Proj3.
 XX KW Erythrocyte binding ligand; Proj3; binding domain; malaria; therapy;
 XX KM vaccine.
 OS Plasmodium falciparum.
 XX PN WO9507353-A.
 PD 16-MAR-1995.
 XX PF 07-SEP-1994; 94WO-US10230.
 XX PR 10-SEP-1993; 93US-0119677.
 (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 DR WPI; 1995-123427/16.
 DR N-PSDB; AA083529.
 XX PT New erythrocyte binding domain polypeptide(s) - isolated from
 PT Plasmodium binding proteins, used in diagnosis, treatment and
 PT prevention of malaria
 XX PS
 PS Disclosure; Page 61-65; 81pp; English.
 CC Erythrocyte binding ligand (EBL) family genes were cloned from
 CC P. falciparum chromosome 7 subsegment libraries constructed during
 CC genetic studies of the chloroquine resistance locus. The 4 genes,
 CC EBL-e1 (AA083526), E31a (AA083527), EBL-e2 (AA083528) and Proj3
 CC (AA083529), encode the proteins given in AAR70233-36, respectively. The
 CC binding domains of such proteins can be expressed e.g. in E. coli,
 CC yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
 CC cells, and provide protection against P. falciparum.
 SQ Sequence 2703 AA;

Query Match 3.6%; Score 10; DB 16; Length 2703;
 Best Local Similarity 100.0%; Pred. No. 0.93;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 28 CMPPRRQKLC 37
 |||||
 Db 1384 CMPPRRQKLC 1393
 RESULT 13
 AAW22482
 ID AAW22482 standard; Protein; 2710 AA.
 XX AC
 XX AAW22482;
 XX DT 07-OCT-1997 (first entry)
 XX DE Plasmodium Proj3.
 XX KM DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
 XX KM Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 XX KM DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KM Plasmodium.
 OS Plasmodium falciparum.
 XX PN WO9640766-A2.
 PD 19-DEC-1996.
 XX PF 07-JUN-1996; 96WO-US09508.
 XX PR 07-JUN-1995; 95US-0487826.
 (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 DR WPI; 1997-052231/05.
 DR N-PSDB; AAT72897.
 XX PT New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and sialic acid
 PT binding proteins
 XX PS
 PS Disclosure; Page 50-56; 96pp; English.
 CC This sequence represents Proj3 of Plasmodium. Proj3 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and sialic acid binding protein
 CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 CC var family of genes modulate cytoadherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and
 CC SABP mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the
 CC invention. The compositions are for the treatment and prevention of
 CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
 CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family. A
 CC family of genes having homology with conserved regions of DABP and SABP.
 CC The compositions are used for the treatment and prevention of malaria.
 CC They are also used in the preparation of vaccines for inducing a
 CC protective immune response in a mammal to Plasmodium merozoites
 CC (especially Plasmodium falciparum or Plasmodium vivax).
 SQ Sequence 2710 AA;
 Query Match 3.6%; Score 10; DB 18; Length 2710;
 Best Local Similarity 100.0%; Pred. No. 0.94;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 CMPRRRQKLC 37
 |||||
 Db 1391 CMPRRRQKLC 1400

RESULT 14
 AAY77904
 ID AAY77904 standard; Protein; 2710 AA.

XX AAY77904;

DT 13-JUN-2000. (first entry)

DE P. falciparum Proj3 binding domain polypeptide.

KM DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
 KM DABP; Stalic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
 KM Protozoacide; Proj3.

US Plasmodium falciparum.

XX US993827-A.

PN 30-NOV-1999.

PD 07-JUN-1995; 95US-0487826.

PF 10-SEP-1993; 93US-0119677.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;

PI WPI; 2000-194198/17.

DR N-PSDB; AA298286.

XX Isolated protein binding domains from Plasmodium vivax and Plasmodium

PT falciparum erythrocyte binding proteins useful for vaccinating against

PT malaria -

XX Disclosure; Columns 79-92; 93pp; English.

PS The invention relates to ebl-1 polypeptides that are encoded by the DBL

CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially

CC identical to the Duffy Antigen Binding Protein (DABP) and Stalic Acid

CC Binding Protein (SABP), which are soluble proteins that appear in the

CC culture supernatant after erythrocytes infected with malaria release

CC merozoites. Immunochemical studies indicate that DABP and SABP are the

CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy

CC and stalic acid receptors on erythrocytes. The ebl-1 polypeptides may be

CC used to vaccinate against malaria, especially caused by P. falciparum.

CC Immunization with the polypeptide provides effective protection against

CC malaria. The present sequence represents the Proj3 binding domain

CC polypeptide.

XX Sequence 2710 AA;

XX Query Match 3.6%; Score 10; DB 21; Length 2710;

XX Best Local Similarity 100.0%; Pred. No. 0.94;

XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 12-SEP-1997 (first entry)

DT Plasmodium var-7.

XX DBL gene family; SABP; stalic acid binding protein; vaccine; therapy;

XX Duffy binding like gene; Duffy antigen binding protein; erythrocyte;

XX DABP; merozoite; malaria; var-1; var-2; var-3; var-7; Immune response;

XX Plasmodium.

XX Plasmodium vivax.

OS Plasmodium falciparum.

XX WO9640766-A2.

XX 19-DEC-1996.

XX 07-JUN-1996; 96WO-US09508.

XX 07-JUN-1995; 95US-0487826.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;

XX Wellens TE;

PI WPI; 1997-052231/05.

DR N-PSDB; AAT72882.

XX New malaria vaccines - contains cysteine-rich DBL family protein

PT binding domains homologous domains of the Duffy and stalic acid

PT binding proteins

XX Claim 8; Page 61-67; 96pp; English.

XX This sequence represents var-7 of Plasmodium. Var-7 belongs to

CC the Duffy binding like (DBL) family of genes which have homology to the

CC Duffy antigen binding protein (DABP) and stalic acid binding protein

CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The

CC var family of genes modulate cytoadherence and antigenic variation of

CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding

CC protein (DABP) are soluble proteins that appear in the culture

CC supernatant after infected erythrocytes release merozoites. DABP and

CC SABP mediate the binding of merozoites and schizonts to the erythrocyte

CC surface. These proteins are necessary for erythrocyte invasion by the

CC parasite. This sequence can be used in the compositions of the

CC invention. The compositions are for the treatment and prevention of

CC malaria, and comprise either a nucleotide sequence or encoded polypeptide

CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a

CC family of genes having homology with conserved regions of DABP and SABP.

CC The compositions are used for the treatment and prevention of malaria.

CC They are also used in the preparation of vaccines for inducing a

CC protective immune response in a mammal to Plasmodium merozoites

CC (especially Plasmodium falciparum or Plasmodium vivax).

XX Sequence 3060 AA;

XX Query Match 3.6%; Score 10; DB 18; Length 3060;

XX Best Local Similarity 100.0%; Pred. No. 1;

XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 CMPRRRQKLC 37
 |||||
 Db 1389 CMPRRRQKLC 1398

Search completed: April 28, 2003, 10:49:46
 Job time : 51.6111 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:48:21 ; Search time 24.8889 Seconds

(without alignments)
1066.062 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554

Perfect score: 276
Sequence: 1 DCHPKKNSNGYDWCQGNIN.....KKEMIPDKYRDERDKRF 276

Scoring table: OLIGO
Gapop 60.0 , Gapept 60.0

Searched: 283224 seqs, 96134422 residues

rd size : 9

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	6.2	3026	2 T28431	variant surface pr
2	11	4.0	3006	2 T28625	variant-specific s
3	10	3.6	1711	2 C71625	variant-specific s
4	10	3.6	3078	2 T28432	variant-specific s

ALIGNMENTS

RESULT 1

T28431 variant surface protein 1 homolog A4VAR - malaria parasite (Plasmodium falciparum) (frag

C:Species: Plasmodium falciparum

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: T28431

R:Smith, J.D.; Kyes, S.; Craig, A.G.; Fegan, T.; Hudson-Taylor, D.; Miller, L.H.; Barruch
Mol. Biochem. Parasitol. 97, 133-148, 1998

A:Title: Analysis of adhesive domains from the A4VAR Plasmodium falciparum erythrocyte m

A:Reference number: 220486; PMID:99094502; PMID:9879893

A:Accession: T28431

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3026 <SMT>

A:Cross-references: EMBL:LA2244; NID:G3540144; PID:G3540145; PIDN:AAD03351.1

C:Genetics:
A:Gene: var

A:Introns: 2906/3

Query Match 6.2%; Score 17; DB 2; Length 3026;

Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 107 GDYRDFLEGTDISKGHG 123
|||||
Db 2122 GDYRDFLEGTDISKGHG 2138

RESULT 2

T28625

variant-specific surface protein 3 - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: T28625

R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Gulnet, F.; Herfelde, J.A.; Peterson,
Cell 82, 89-100, 1995

A:Title: The large diverse gene family var encodes proteins involved in cytoadherence

A:Reference number: 220487; PMID:95330813; PMID:7606788

A:Accession: T28625

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3006 <SU>

A:Cross-references: EMBL:LA0609; NID:G886376; PID:G886377; PIDN:AA5397.1

C:Genetics:

A:Gene: var-3

A:Introns: 2597/3

Query Match

Best Local Similarity 100.0%; Pred. No. 0.022;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 CMPPRKRLCV 38
|||||

Db 1270 CMPPRKRLCV 1280

RESULT 3

C71625

variant-specific surface protein 1 homolog PFB0010w - malaria parasite (Plasmodium fa

C:Species: Plasmodium falciparum

C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000

C:Accession: C71625

R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
Science 281, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; PMID:99021743; PMID:9804551

A:Accession: C71625

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1711 <GAR>

A:Cross-references: GB:AE001366; GB:AE001362; NID:G3845070; PIDN:AMC71792.1; PID:G384

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0010w

Query Match

Best Local Similarity 100.0%; Pred. No. 0.15;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 MFYFGDYRD 111
|||||

Db 951 MFYFGDYRD 960

RESULT 4

T28432

variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)

N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)

C:Species: Plasmodium falciparum

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: T28432

R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Gulnet, F.; Herfelde, J.A.; Peterson,

Cell 82, 89-100, 1995
 A>Title: The large diverse gene family var encodes proteins involved in cytoadherence at
 A:Reference number: Z20487; MID:95330813; PMID:7606788
 A:Accession: T28432
 A>Status: preliminary; translated from GR/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3078 <STX>
 A:Cross-references: EMBL:L40608; NID:9886374; PID:9886375; PIDN:AAA75396.1
 C:Genetics:
 A:Gene: var-1
 A:introns: 2611/3

Query Match 3.68; Score 10; DB 2; Length 3078;
 Best Local Similarity 100.0%; Pred. No. 0.25;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 28 CMPPRQKLC 37
 |||||
 Db 1389 CMPPRQKLC 1398

rch completed: April 28, 2003, 10:52:29
 time : 26.8889 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:35:40 ; Search time 14.2222 seconds
(without alignments)
804.900 Million cell updates/sec

Title: us-10-087-013-2_COPY_1279_1554

Perfect score: 276
Sequence: 1 DCHPRKNSNGYPPDMQGNIN.....KKEWIPQDKRYKDERDKRRF 276

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Hit size : 9

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

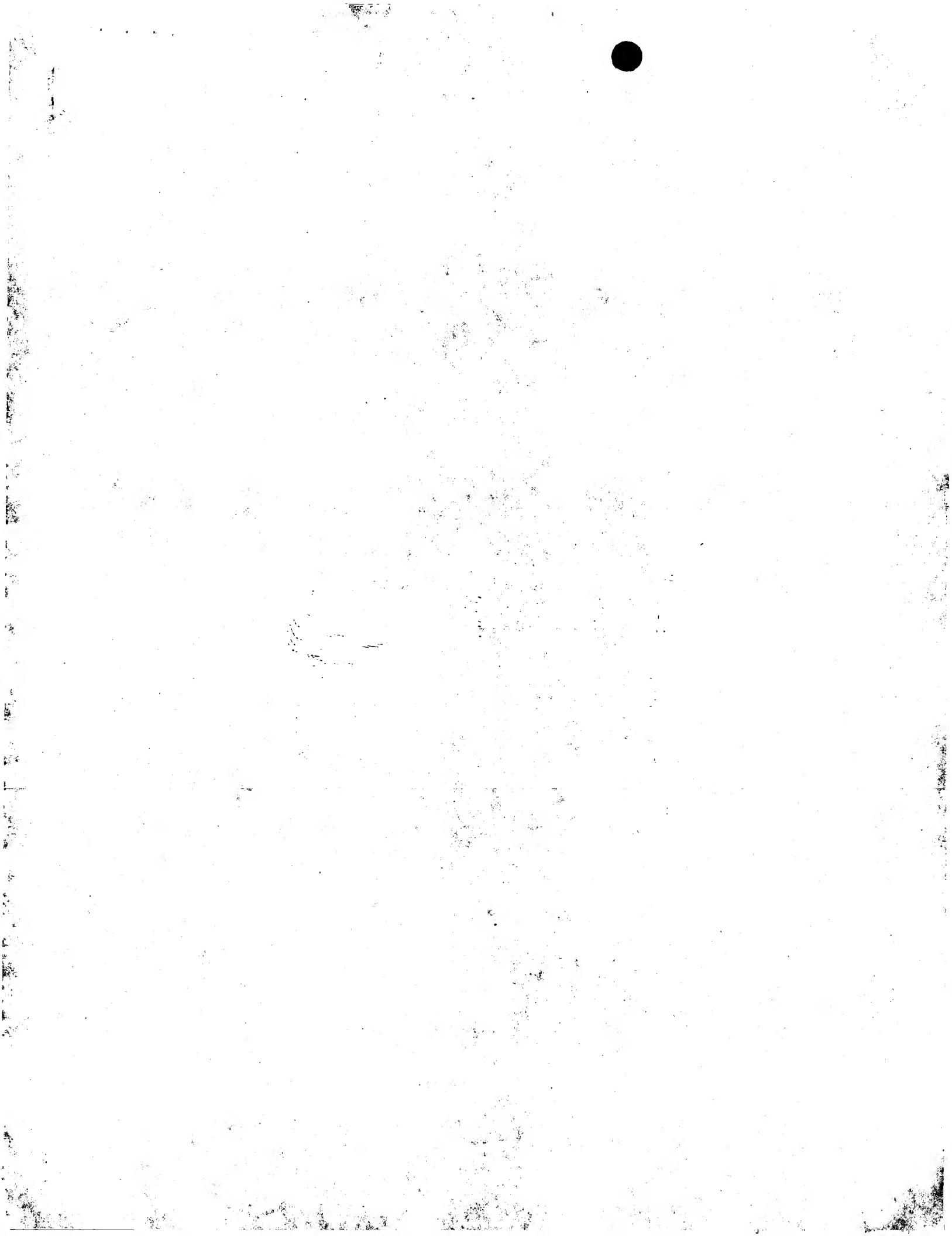
Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	ID	Description
No matches found			

Search completed: April 28, 2003, 10:50:21
Job time : 14.2222 secs



GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:45:26 ; Search time 25.333 Seconds

(without alignments)
2244.829 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554

Perfect score: 276
Sequence: 1 DCHPRKNSNGYPMQCGNIN.....KKEMIPQDKYKDERDKRRF 276

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

rd size : 9

Total number of hits satisfying chosen parameters: 19

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- 1: SP-archaea:*
- 2: SP-bacteria:*
- 3: SP-fungi:*
- 4: SP-human:*
- 5: SP-invertebrate:*
- 6: SP-mammal:*
- 7: SP-mhc:*
- 8: SP-organelle:*
- 9: SP-phage:*
- 10: SP-plant:*
- 11: SP-rodent:*
- 12: SP-virus:*
- 13: SP-vertebrate:*
- 14: SP-unclassified:*
- 15: SP-virus:*
- 16: SP-bacteriaph:*
- 17: SP-archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	276	100.0	3542	09U5M2	09U5M2 plasmodium
2	20	7.2	1615	08T325	08T325 plasmodium
3	20	7.2	3287	08T326	08T326 plasmodium
4	17	6.2	169	025990	025990 plasmodium
5	17	6.2	182	000832	000832 plasmodium
6	17	6.2	3026	026030	026030 plasmodium
7	14	5.1	173	025988	025988 plasmodium
8	12	4.3	1685	09U4A2	09U4A2 plasmodium
9	11	4.0	162	09B1E7	09B1E7 plasmodium
10	11	4.0	195	000831	000831 plasmodium
11	11	4.0	921	025989	025989 plasmodium
12	11	4.0	2527	095W83	095W83 plasmodium
13	11	4.0	2710	09XZB8	09XZB8 plasmodium
14	11	4.0	3006	026032	026032 plasmodium
15	10	3.6	81	09NFB4	09NFB4 plasmodium
16	10	3.6	1711	096108	096108 plasmodium

ALIGNMENTS

17	10	3.6	2658	5	08T5G0	08T5G0 plasmodium
18	10	3.6	3078	5	Q26031	Q26031 plasmodium
19	9	3.3	1327	5	Q9NFB4	Q9NFB4 plasmodium

RESULT 1

ID	Score	Query Match	Length	ID	Description
09U5M2	276	100.0	3542	09U5M2	09U5M2 plasmodium
AC	09U5M2	PRELIMINARY;	PRT;	3542	AA.
DT	01-MAY-2000	(TREMblrel. 13, Created)			
DT	01-MAY-2000	(TREMblrel. 13, Last sequence update)			
DT	01-MAR-2002	(TREMblrel. 20, Last annotation update)			
DE	FCR3	CSA ligand (Fragment).			
GN	VAR.				
OS	Plasmodium	falciparum.			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxID=5633;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=FCR3;				
RX	MEDLINE=20006305; PubMed=10535993;				
RA	Buffet P., Gamain B., Scheldis C., Baruch B., Oishi S., Fujii N.,				
RA	Fusai T., Parzy D., Miller L.H., Gysin J., Schert A.;				
RT	"Plasmodium falciparum domain mediating adhesion to Chondroitin				
RT	sulfate A: A receptor for human placental infection."				
RL	Proc. Natl. Acad. Sci. U.S.A. 96:12743-12748(1999).				
DR	EMBL: AJ133811; CAB59840.1;				
DR	InterPro: IPR001219; Neurotoxin.				
DR	InterPro: IPR004258; PFEMP.				
DR	Pfam: PF03011; PFEMP. 1.				
DR	PRINTS: PRO0284; TOXIN.				
FP	NON TER. 3542 3542				
SO	SEQUENCE 3542 AA; 413089 MW; 970D85E86DA2EC2 CRC64;				

Query Match 100.0%; Score 276; DB 5; Length 3542;
Best Local Similarity 100.0%; Pred. No. 2.9e-284;
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	DCHPRKNSNGYPMQCGNINLYEDPRVCMPPRRQCLCYHFLANDNEIKKLSQVNLKNAF	60
DB	1279	DCHPRKNSNGYPMQCGNINLYEDPRVCMPPRRQCLCYHFLANDNEIKKLSQVNLKNAF	1338
QY	61	IKSAAETFFSWYKSKDGEENELDKELKEKIPAFILRSWYFGDYRDLFTGDTISK	120
DB	1339	IKSAAETFFSWYKSKDGEENELDKELKEKIPAFILRSWYFGDYRDLFTGDTISK	1398
QY	121	GHGEGSKLEQIDSLFKNGDQSPNGKTRQEMWTEHSHEIWEAMICALVKGAKKDFTE	180
DB	1399	GHGEGSKLEQIDSLFKNGDQSPNGKTRQEMWTEHSHEIWEAMICALVKGAKKDFTE	1458
QY	181	NYGVNNVKSFSKSTLEEFARPOFLRLTEYDYCYTRQKYLTDVQKCSNOIKCD	240
DB	1459	NYGVNNVKSFSKSTLEEFARPOFLRLTEYDYCYTRQKYLTDVQKCSNOIKCD	1518
QY	241	TECNKCEDEYKYMKKKEMIPQDKYKDERDKRRF	276
DB	1519	TECNKCEDEYKYMKKKEMIPQDKYKDERDKRRF	1554
RESULT 2			
08T325			
ID	08T325	PRELIMINARY;	PRT; 1615 AA.
AC	08T325		
DT	01-JUN-2002	(TREMblrel. 21, Created)	
DT	01-JUN-2002	(TREMblrel. 21, Last sequence update)	
DT	01-JUN-2002	(TREMblrel. 21, Last annotation update)	
DE	PFEMP1	(Fragment).	
GN	TM284VAR3.		
OS	Plasmodium	falciparum.	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		

OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TM284;
RX MEDLINE=21927235; PubMed=11930336;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
RL J. Infect. Dis. 185:1207-1211(2002).
DR EMBL; AJ420412; CAD20868.1; -
FT NON_TER 1615 1615
SQ SEQUENCE 1615 AA; 188639 MW; FCA896C00B86DEA6 CRC64;

Query Match 7.2%; Score 20; DB 5; Length 1615;
Best Local Similarity 100.0%; Pred. No. 5.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 CKSNDOLKCDTECNKCCDY 250
Db 1495 CKSNDOLKCDTECNKCCDY 1514
|||||
SOUT 3
ID 08T326 PRELIMINARY; PRT; 3287 AA.
AC 08T326;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE PLEMP1 (Fragment).
GN TM180VAR2.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TM180;
RX MEDLINE=21927235; PubMed=11930336;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
RL J. Infect. Dis. 185:1207-1211(2002).
DR EMBL; AJ420411; CAD20867.1; -
FT NON_TER 3287 3287
SQ SEQUENCE 3287 AA; 383550 MW; 58F8F866FC244536 CRC64;

Query Match 7.2%; Score 20; DB 5; Length 3287;
Best Local Similarity 100.0%; Pred. No. 9.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

231 CKSNDOLKCDTECNKCCDY 250
Db 1495 CKSNDOLKCDTECNKCCDY 1514
|||||

RESULT 4
ID 025990 PRELIMINARY; PRT; 169 AA.
AC 025990;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE (clone pNM4) ORF (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DD2/NM;
RX MEDLINE=95350219; PubMed=7624377;
RA Peterson D.S., Miller L.H., Wellens T.E.;
RT "Isolation of multiple sequences from the Plasmodium falciparum genome
that encode conserved domains homologous to those in erythrocyte-

RT binding proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:7100-7104(1995).
DR EMBL; L38455; AAC37241.1; -
FT NON_TER 1 1
SQ SEQUENCE 169 AA; 19120 MW; BCC6B0FC3FE1380 CRC64;

Query Match 6.2%; Score 17; DB 5; Length 169;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 GDYRDFLEFGTDSKGHG 123
Db 71 GDYRDFLEFGTDSKGHG 87
|||||

RESULT 5
ID 000832 PRELIMINARY; PRT; 182 AA.
AC 000832;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Partial erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum (Isolate Dd2).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=57267;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DD2;
RX MEDLINE=97342604; PubMed=9199301;
RA Fischer K., Horrocks P., Preuss M., Wiesner J., Wensch S.,
RA Camargo A.A., Lanzer M.;
RT "Expression of var genes located within polymorphic subtelomeric
domains of Plasmodium falciparum chromosomes.";
RL Mol. Cell. Biol. 17:3679-3686(1997).
DR EMBL; Y11910; CAA72661.1; -
FT NON_TER 1 1
SQ SEQUENCE 182 AA; 20830 MW; 197CA46778AD148B CRC64;

Query Match 6.2%; Score 17; DB 5; Length 182;
Best Local Similarity 100.0%; Pred. No. 1.2e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 GDYRDFLEFGTDSKGHG 123
Db 78 GDYRDFLEFGTDSKGHG 94
|||||

RESULT 6
ID 026030 PRELIMINARY; PRT; 3026 AA.
AC 026030;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Variant surface protein (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE OF 2044-2922 FROM N.A.
RC STRAIN-IT 4/25/5;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herfeldt J.A.,
RA Peterson D.S., Ravetch J.A., Wellens T.E.;
RT "The large diverse gene family var encodes proteins involved in
cytoadherence and antigenic variation of Plasmodium falciparum-
infected erythrocytes.";
RL Cell 82:89-100(1995).

[2]
RN SEQUENCE OF 2044-2922 FROM N.A.
RP STRAIN-IT 4/25/5;
RA Smith J.D., Chitnis C.E., Craig A.G., Roberts D.J.,
RA Hudson-Taylor D.E., Peterson D.S., Pinches R., Newbold C.I.,
RA Miller L.H.;
RT "Switches in the expression of Plasmodium falciparum var genes
correlate with changes in antigenic and cytoadherent phenotypes of
infected erythrocytes.";
RT Cell 0:0-0(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-IT 4/25/5;
RA MEDLINE-99094502;
RA Smith J.D., Kyes S., Craig A.G., Fagan T., Hudson-Taylor D.,
RA Miller L.H., Baruch D.I., Newbold C.I.;
RT "Analysis of adhesive domains from the AAVAR Plasmodium falciparum
erythrocyte membrane protein-1 identifies a CD36 binding domain.";
RT Mol. Biochem. Parasitol. 97:133-148(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-IT 4/25/5;
RA Smith J.;
RT Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-IT 4/25/5;
RA Kyes S., Smith J.;
RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN EMBL: L42244; AAD03351.1; -;
DR InterPro: IPR004256; PFEMP.
DR Pfam: PF03011; PFEMP; 2.
FT NON_TER 3026 3026
SQ SEQUENCE 3026 AA; 348124 MW; EC0A971A82193788 CRC64;

Query Match 6.28; Score 17; DB 5; Length 3026;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 GDYRDLFGTIDSKGKG 123
Db 2122 GDYRDLFGTIDSKGKG 2138
|||||

RESULT 7
Q25988 PRELIMINARY; PRT; 173 AA.
ID Q25988;
AC 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DE (clone pM5) ORF (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DD2/NM;
RA MEDLINE-95350219; Pubmed-7624377;
RA Peterson D.S., Miller L.H., Wellens T.E.;
RT "Isolation of multiple sequences from the Plasmodium falciparum genome
that encode conserved domains homologous to those in erythrocyte-
binding proteins.";
RT Proc. Natl. Acad. Sci. U.S.A. 92:7100-7104(1995).
DR EMBL: L38453; AAC37239.1; -;
FT NON_TER 1 1
FT NON_TER 173 173
SQ SEQUENCE 173 AA; 19828 MW; 365051FBF25ED2P2 CRC64;

Query Match 5.18; Score 14; DB 5; Length 173;
Best Local Similarity 100.0%; Pred. No. 1.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 RDLFGTIDSKGKG 123
Db 70 RDLFGTIDSKGKG 83
|||||

RESULT 8
ID Q904A2 PRELIMINARY; PRT; 1685 AA.
AC Q904A2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
DE 01-MAR-2002 (TREMBlrel. 20, last annotation update)
DE Variant surface protein PfEMP1 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IT;
RA MEDLINE-20144115; Pubmed-10677532;
RA Smith J.D., Craig A.G., Kriek N., Hudson-Taylor D., Kyes S., Fagan T.,
RA Pinches R., Baruch D.I., Newbold C.I., Miller L.H.;
RT "Identification of a Plasmodium falciparum intercellular adhesion
molecule-1 binding domain: A parasite adhesion trait implicated in
cerebral malaria.";
RT Proc. Natl. Acad. Sci. U.S.A. 97:1766-1771(2000).
DR EMBL: AF193424; AAF18980.1; -;
DR InterPro: IPR000454; Eub_ATPase_Csub.
DR InterPro: IPR004256; PFEMP.
DR Pfam: PF03011; PFEMP; 1.
DR PROSITE: PS00605; ATPASE_C; 1.
FT NON_TER 1685 1685
SQ SEQUENCE 1685 AA; 193671 MW; D1FD42666B0551E CRC64;

Query Match 4.38; Score 12; DB 5; Length 1685;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 EAFKSAAEET 69
Db 1293 EAFKSAAEET 1304
|||||

RESULT 9
ID Q9BJE7 PRELIMINARY; PRT; 162 AA.
AC Q9BJE7;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-720;
RA MEDLINE-21136462; Pubmed-11237850;
RA Khattab A., Kun J., Deloron P., Krensner P.G., Klinkert M.O.;
RT "Variants of Plasmodium falciparum Erythrocyte Membrane Protein 1
Expressed by Different Placental Parasites are Closely Related and
Adhere to Chondroitin Sulfate A.";
RT J. Infect. Dis. 183:1165-1169(2001).
DR EMBL: AF334806; AAK28129.1; -;
FT NON_TER 1 1
FT NON_TER 162 162
SQ SEQUENCE 162 AA; 18788 MW; BC2E24182BF1078E CRC64;

Query Match 4.08; Score 11; DB 5; Length 162;
Best Local Similarity 100.0%; Pred. No. 0.0026;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 AFKSAAEFT 69
 DB 25 AFKSAAEFT 35

RESULT 10

000831 PRELIMINARY; PRT; 195 AA.
 AC 000831;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 GN Partial erythrocyte membrane protein 1 (Fragment).
 OS Plasmodium falciparum (isolate Dd2).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=57267;
 RN [1]
 RC STRAIN-DD2;
 MEDLINE=97342604; PubMed=9199301;
 Fischer K., Horrocks P., Preuss M., Wiesner J., Wensch S.,
 Camargo A.A., Lanzer M.;
 "Expression of var genes located within polymorphic subtelomeric
 domains of Plasmodium falciparum chromosomes.";
 RT Mol. Cell. Biol. 17:3679-3686(1997).
 DR EMBL; Y11909; CAA72660.1; -.
 FT NON_TER 1
 SQ SEQUENCE 195 AA; 22667 MW; 8B5F1D5F18366F21 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 11; DB 5; Length 195;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 110 RDEFGTDISK 120
 DB 88 RDEFGTDISK 98

RESULT 11

025989 PRELIMINARY; PRT; 921 AA.
 AC 025989;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE (clone pS31H) ORF (Fragment).
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX STRAIN-DD2/M;
 MEDLINE=95350219; PubMed=7624377;
 RA Peterson D.S., Miller L.H., Wellens T.E.;
 "Isolation of multiple sequences from the Plasmodium falciparum genome
 that encode conserved domains homologous to those in erythrocyte-
 binding proteins.";
 RT Proc. Natl. Acad. Sci. U.S.A. 92:7100-7104(1995).
 RU EMBL; J38454; AAC37240.1; -.
 DR EMBL; J38454; AAC37240.1; -.
 DR InterPro; IPR004258; PFEMP.
 DR Pfam; PF03011; PFEMP; 1.
 FT NON_TER 1
 SQ SEQUENCE 921 AA; 104711 MW; D98BC07763628746 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 11; DB 5; Length 921;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 202 RPOFLRWLEW 212
 DB 618 RPOFLRWLEW 628

RESULT 12

095W83 PRELIMINARY; PRT; 2527 AA.
 AC 095W83;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Erythrocyte membrane protein 1 (Fragment).
 GN VAR1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX MEDLINE=21442075; PubMed=11557894;
 RA Flick K., Scholander C., Chen Q., Fernandez V., Pouvelle B., Gysin J.,
 Wahlgren M.;
 "Role of nonimmune IgG bound to PfEMP1 in placental malaria.";
 RT Science 293:2098-2100(2001).
 DR EMBL; AF366567; AAL12845.1; -.
 DR InterPro; IPR004258; PFEMP.
 DR Pfam; PF03011; PFEMP; 2.
 FT NON_TER 2527
 SQ SEQUENCE 2527 AA; 292137 MW; 59E2766BF0425E4 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 11; DB 5; Length 2527;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 AFKSAAEFT 69
 DB 1315 AFKSAAEFT 1325

RESULT 13

09XZB8 PRELIMINARY; PRT; 2710 AA.
 AC 09XZB8;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE Variant-specific surface protein.
 GN VAR.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX STRAIN-CS2;
 MEDLINE=99238507; PubMed=10220443;
 RA Reeder J.C., Cowman A.F., Davern K.M., Beeson J.G., Thompson J.K.,
 Rogerson S.J., Brown G.V.;
 "The adhesion of Plasmodium falciparum-infected erythrocytes to
 chondroitin sulfate A is mediated by P. falciparum erythrocyte
 membrane protein 1.";
 RT Proc. Natl. Acad. Sci. U.S.A. 96:5198-5202(1999).
 RU EMBL; AF134154; AAD29126.1; -.
 DR EMBL; AF134154; AAD29126.1; -.
 DR InterPro; IPR004258; PFEMP.
 DR Pfam; PF03011; PFEMP; 2.
 FT NON_TER 2
 SQ SEQUENCE 2710 AA; 309247 MW; 0AC64F0A3D5BF512 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 11; DB 5; Length 2710;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 CMPPRKLCV 38
 DB 941 CMPPRKLCV 951

RESULT 14

026032
 ID 026032 PRELIMINARY; PRT: 3006 AA.
 AC 026032;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE Variant-specific surface protein.
 GN VAR-3.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FCR3;
 RX MEDLINE=95330813; PubMed=7606788;
 RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
 RA Peterson D.S., Ravetch J.A., Wellem T.E.;
 RT "The large diverse gene family var encodes proteins involved in
 cytoadherence and antigenic variation of Plasmodium falciparum-
 infected erythrocytes."
 Cell 82:89-100(1995).
 DR EMBL; L40609; AAA75397.1; -;
 DR InterPro: IPR004258; PfEMP.
 DR Pfam: PF03011; PfEMP; 2
 SO SEQUENCE 3006 AA; 343769 MW; 8FD5F475F943C7A CRC64;

Query Match 4.0%; Score 11; DB 5; Length 3006;
 Best Local Similarity 100.0%; Pred. No. 0.034;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 CMPRRKRLCV 38
 |||||
 DB 1270 CMPRRKRLCV 1280

RESULT 15
 Q9NFM8 PRELIMINARY; PRT: 81 AA.
 ID Q9NFM8
 AC Q9NFM8;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE Var protein (Fragment).
 GN VAR.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21316723; PubMed=11424024;
 RA Arley F., Hommel D., Le Scanf C., Duchemin J.B., Peneau C., Hulin A.,
 RA Sarthou J.L., Reyes J.M., Fandeur T., Mercereau-Pujalon O.;
 RT "Association of severe malaria with a specific Plasmodium falciparum
 RT genotype in French Guiana."
 J. Infect. Dis. 184:237-241(2001).
 DR EMBL; AJ27137; CAB86464.1; -;
 KW Malaria.
 FT NON_TER 1
 FT VARIANT 11 15 TFGDY -> IWRH.
 FT NON_TER 81
 FT NON_TER 81
 SO SEQUENCE 81 AA; 9336 MW; 5849685D4A336C7A CRC64;

Query Match 3.6%; Score 10; DB 5; Length 81;
 Best Local Similarity 100.0%; Pred. No. 0.016;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 102 MFTFGDYRD 111
 |||||
 DB 8 MFTFGDYRD 17

Search completed: April 28, 2003, 10:51:30
 Job time : 29.333 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:49:21 ; Search time 16.8889 Seconds
(without alignments)
480.832 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554

Perfect score: 276
Sequence: 1 DCHPKKNSNGYPMQCGNIN.....KKEMIPQDKYKDERDKRRP 276

Scoring table: OLIGO
Gapop 60.0 , Gapect 60.0

Searched: 262574 seqs, 29422922 residues

rd size: 9

Total number of hits satisfying chosen parameters: 13

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	4.0	324	2 US-08-568-459A-17	Sequence 17, Appl
2	11	4.0	324	2 US-08-487-826B-29	Sequence 29, Appl
3	11	4.0	324	4 US-09-210-288-17	Sequence 17, Appl
4	11	4.0	921	2 US-08-568-459A-8	Sequence 8, Appl
5	11	4.0	921	2 US-08-487-826B-8	Sequence 8, Appl
6	11	4.0	921	2 US-09-210-288-8	Sequence 8, Appl
7	10	3.6	311	2 US-08-568-459A-21	Sequence 21, Appl
8	10	3.6	311	4 US-08-487-826B-33	Sequence 33, Appl
9	10	3.6	311	4 US-09-210-288-21	Sequence 21, Appl
10	10	3.6	2710	2 US-08-568-459A-12	Sequence 12, Appl
11	10	3.6	2710	2 US-08-487-826B-12	Sequence 12, Appl
12	10	3.6	2710	4 US-09-210-288-12	Sequence 12, Appl
13	10	3.6	3060	2 US-08-487-826B-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-568-459A-17
; Sequence 17, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.

APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-0176
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
US-08-568-459A-17

Query Match 4.0%; Score 11; DB 2; Length 324;
Best Local Similarity 100.0%; Pred. No. 0.0034;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 202 RPOFLRWLEW 212
Db 204 RPOFLRWLEW 214

RESULT 2
US-08-487-826B-29
; Sequence 29, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
US-08-487-826B-29

Query Match
Best Local Similarity 4.0%; Score 11; DB 2; Length 324;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 202 RPOFLRWLTW 212
|||||
DB 204 RPOFLRWLTW 214

RESULT 3
US-09-210-288-17
Sequence 17, Application US/09210288
Patent No. 6392026
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
US-09-210-288-17

Query Match
Best Local Similarity 4.0%; Score 11; DB 4; Length 324;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 202 RPOFLRWLTW 212
|||||
DB 204 RPOFLRWLTW 214

RESULT 4
US-08-568-459A-8
Sequence 8, Application US/08568459A
Patent No. 5849306

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-8

Query Match
Best Local Similarity 4.0%; Score 11; DB 2; Length 921;
Matches 100.0%; Pred. No. 0.0083;

	Matches	11; Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	202	RQFRLRWLEW	212						
Db	618	RQFRLRWLEW	628						

RESULT 5
 US-08-487-826B-8
 : Sequence 8, Application US/08487826B
 : Patent No. 5993827
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Sim, Kim L.
 : APPLICANT: Chlenuis, Chetan
 : APPLICANT: Miller, Louis H.
 : APPLICANT: Peterson, David S.
 : APPLICANT: Su, Xin-zhaun
 :
 : APPLICANT: Wellens, Thomas E.
 :
 : TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 :
 : NUMBER OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 :
 : NUMBER OF SEQUENCES: 45
 :
 : CORRESPONDENCE ADDRESS:

```

? MOLECULE TYPE: Protein
? HYPOTHETICAL: NO
? ORIGINAL SOURCE:
? ORGANISM: Plasmodium falciparum
US-08-487-826B-8

Query Match      4.0%  Score 11;  DB 2;  Length 921;
Best Local Similarity 100.0%  Pred. No. 0.0083;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0

```

RESULT 6
US-09-210-288-8
Sequence 8, Application US/092102886
Patent No. 6392026
GENERAL INFORMATION:
APPLICANT: Sim, Kam L.,
APPLICANT: Chinnis, Chetan
APPLICANT: Miller, Louis H.

```

1  APPLICANT: Peterson, David S.
2  APPLICANT: Su, Xin-zhaun
3  APPLICANT: Wellens, Thomas E.
4  TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
5  TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
6  NUMBER OF SEQUENCES: 37
7  CORRESPONDENCE ADDRESS:
8  ADDRESSEE: Knobbe Martens Olson & Bear
9  STREET: 620 Newport Center Drive 16th Floor
10 CITY: Newport Beach
11 STATE: California
12 COUNTRY: US
13 ZIP: 92660
14
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: PatentIn Release #1.0, Version #1.25
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/09/210,288
22 FILING DATE:
23
24 CLASSIFICATION:
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Fuller, Michael
27 REGISTRATION NUMBER: 36,516
28 REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: (619) 235-8550
31 TELEFAX: (619) 235-0176
32
33 INFORMATION FOR SEQ ID NO: 8:
34 SEQUENCE CHARACTERISTICS:
35     LENGTH: 921 amino acids
36     TYPE: amino acid
37     STRANDEDNESS: single
38     TOPOLOGY: linear
39     MOLECULE TYPE: protein
40     HYPOTHETICAL: NO
41     ORIGINAL SOURCE:
42     ORGANISM: Plasmodium falciparum
43
44 US-09-210-288-8
45
46 Query Match 4.0%; Score 11; DB 4; Length 921;
47 Best Local Similarity 100.0%; Pred. No. 0.0083;
48 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
49
50 09 202 RPOFLRWLTW 212
51 DB 618 RPOFLRWLTW 628

```

RESULT 7
 US-08-568-459A-21
 : Sequence 21, Application US/08568459A
 : Patent No. 5849306
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Sim, Kim L.
 : APPLICANT: Chitnis, Chetan
 : APPLICANT: Miller, Louis H.
 : APPLICANT: Peterson, David S.
 : APPLICANT: Su, Xin-zhaun
 : APPLICANT: Wellens, Thomas E.
 :
 : TITLE OF INVENTION: BINDING DOMAINS FROM PLASMIDIUM VIVAX
 : TITLE OF INVENTION: AND PLASMIDIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 :
 : NUMBER OF SEQUENCES: 37
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Knobbe Martens Olson & Bear
 : STREET: 620 Newport Center Drive
 : CITY: Newport Beach
 : STATE: California
 : COUNTRY: US
 : ZIP: 92660
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk

ADDRESS: Knudde Martins Olson & Bear
STREET: 620 Newport Center Drive
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DAT:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israel, Isten, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NHI21.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176

```

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: N1121.1EWDVA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 311 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEetical: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
US-09-210-288-21

```

Query Match 3.68; Score 10; DB 4; Length 311;

Best Local Similarity 100.0%; Pred. No. 0.033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 28 CMPPRKLC 37
Db 24 CMPPRKLC 33

RESULT 10
US-08-568-459A-12
Sequence 12, Application US/08568459A
Patent No. 5849306

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Willems, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO

ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-12

Query Match 3.6%; Score 10; DB 2; Length 2710;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 28 CMPPRKLC 37
Db 1391 CMPPRKLC 1400

RESULT 11
US-08-487-826B-12
Sequence 12, Application US/08487826B
Patent No. 5993827

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan

APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.

APPLICANT: Su, Xin-zhaun
APPLICANT: Willems, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO

ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-487-826B-12

Query Match 3.6%; Score 10; DB 2; Length 2710;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 28 CMPPRKLC 37
Db 1391 CMPPRKLC 1400

RESULT 12
US-09-210-288-12
Sequence 12, Application US/09210288
Patent No. 6392026

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Willems, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:


```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,288
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-0176
; TELEFAX: (619) 235-8550
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-09-210-288-12

```

```

Query Match          3.6%; Score 10; DB 4; Length 2710;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 28 CMPPRRQKLC 37
   |||||
Db 1391 CMPPRRQKLC 1400

```

```

RESULT 13
US-08-487-826B-14
; Sequence 14, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176

```

```

; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3060 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-487-826B-14

```

```

Query Match          3.6%; Score 10; DB 2; Length 3060;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 28 CMPPRRQKLC 37
   |||||
Db 1389 CMPPRRQKLC 1398

```

```

Search completed: April 28, 2003, 10:53:15
Job time : 19.8889 secs

```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Computron Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:51:36 ; Search time 13.333 Seconds
(without alignments)
1658.687 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554

Perfect score: 276
Sequence: 1 DCHPRKNSNGYPMQCNIN.....KKEWIDQKYYKDERDKRF 276

Scoring table: OLIGO
Gapop 60.0 , Gapept 60.0

Searched: 301932 seqs, 80129803 residues

rd size : 9

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	4.0	324	9	US-10-153-273-17
2	11	4.0	921	9	US-10-153-273-8
3	10	3.6	311	9	US-10-153-273-21
4	10	3.6	2710	9	US-10-153-273-12

ALIGNMENTS

RESULT 1
US-10-153-273-17

Sequence 17, Application US/10153273
Patent No. US20020169305A1

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun

Willems, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
STRANDEDNESS: single
MOLECULE TYPE: peptide
TOPOLGY: linear
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-153-273-17

Query Match 4.0%; Score 11; DB 9; Length 324;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 202 RPOFLRWLTW 212

Db 204 RPOFLRWLTW 214

RESULT 2
US-10-153-273-8

Sequence 8, Application US/10153273
Patent No. US20020169305A1

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Willems, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US

ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-153-273-8

Query Match 4.0%; Score 11; DB 9; Length 921;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 RPOFLRWLTEM 212
Db 618 RPOFLRWLTEM 628

RESULT 3
US-10-153-273-21
Sequence 21, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 311 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-10-153-273-21

Query Match 3.6%; Score 10; DB 9; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 CMPPRQKLC 37
Db 24 CMPPRQKLC 33

RESULT 4
US-10-153-273-12
Sequence 12, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium falciparum
 SEQUENCE DESCRIPTION: SEQ ID NO: 12:
 US-10-153-273-12

Query Match

Best Local Similarity 3.68; Score 10; DB 9; Length 2710;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 CMPPRORKIC 37
 |||||
 Db 1391 CMPPRORKIC 1400

Search completed: April 28, 2003, 11:02:44
 Job time : 14.333 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:46 ; Search time 10.5249 Seconds

(without alignments)
4367.892 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747

Perfect score: 1938
Sequence: 1 TKEKQSTLSNDNKFVNIN.....CKDNNTEACEFSHATTNP 345

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Minimum number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1918	100.0	3542	22	AA62142
2	446	22.3	2228	20	AAW93944
3	432.5	22.5	2197	21	AA18352
4	431	22.5	2445	22	AA66344
5	394.5	20.6	2703	16	AA70236
6	394.5	20.6	2710	18	AAW2248
7	394.5	20.6	2710	21	AA77904
8	394.5	20.6	3060	18	AAW22475
9	394.5	20.6	3060	21	AA77905
10	380	19.8	700	16	AA70235

11	380	19.8	700	18	AAW22481	Plasmodium ebl-2.
12	380	19.8	700	21	AA77903	P. falciparum ebl-
13	380	19.8	2182	18	AAW22476	Plasmodium var-1.
14	380	19.8	2182	21	AA77906	Plasmodium var-1 P
15	357.5	18.6	1726	17	AAW00385	Truncated Plasmodi
16	357.5	18.6	2913	17	AAW00384	Plasmodium falcipa
17	231	12.0	793	16	AA70234	P. falciparum E31a
18	231	12.0	921	18	AAW22480	Plasmodium E31a.
19	231	12.0	921	21	AA77902	P. falciparum ebl-
20	162	8.4	921	21	AAW18311	Plasmodium falcipa
21	161	8.4	972	23	AAU76761	Plasmodium falcipa
22	153.5	8.0	696	21	AAW18181	Plasmodium falcipa
23	151.5	7.9	1086	23	AAU76760	Plasmodium falcipa
24	150.5	7.8	2380	21	AAW18315	Plasmodium falcipa
25	147.5	7.7	1516	21	AAW18195	Plasmodium falcipa
26	147	7.7	1421	23	AAU76764	Plasmodium falcipa
27	145	7.6	749	16	AAW70233	P. falciparum EBL-
28	145	7.6	749	18	AAW22479	Plasmodium ebl-1.
29	145	7.6	749	21	AA77901	P. falciparum ebl-
30	144.5	7.5	1979	21	AAW18171	Plasmodium falcipa
31	143.5	7.5	1166	21	AAW18268	Plasmodium falcipa
32	142.5	7.4	1247	21	AAW18215	Plasmodium falcipa
33	142	7.4	3973	21	AAW18253	Plasmodium falcipa
34	140.5	7.3	1817	21	AAW18255	Plasmodium falcipa
35	137	7.1	1435	16	AAW70232	P. falciparum SAPP
36	137	7.1	1435	18	AAW22477	Silastic acid bindin
37	137	7.1	1435	21	AAW77900	P. falciparum SAPP
38	137	7.1	1604	16	AAW70105	TNF-R-EBA 175 fus1
39	137	7.1	1786	14	AAW41043	CD4-EBA175 fusion
40	137	7.1	2010	21	AAW18218	Plasmodium falcipa
41	136	7.1	508	21	AAW18187	Plasmodium falcipa
42	136	7.1	1147	14	AAW41199	CAI antigen. Heli
43	136	7.1	1338	19	AAW80600	Helicobacter pylor
44	135.5	7.1	350	23	AAW80656	P. falciparum EBA-
45	135.5	7.1	1817	21	AAW18301	Plasmodium falcipa

ALIGNMENTS

RESULT 1
ID AAB62142 standard; Protein: 3542-AA.
XX AAB62142:
AC AAB62142:
XX 29-MAY-2001 (first entry)
DT XX
DE P. falciparum FCR3, varCSA protein.
XX FCR3, varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
KW erythrocyte membrane protein 1; parasitized red blood cell; PBBC;
KW malaria; protozoacide.
XX
OS Plasmodium falciparum.
PN WO200116326-A2.
PN
PD 08-MAR-2001.
PD
PE 01-SEP-2000; 2000WO-US24195.
PE
PR 01-SEP-1999; 99US-0152023.
PR
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
XX Gysin J, Pouvelle B, Fujii N, Smith J;
XX WPI; 2001-235109/24.
XX N-PSDB; AAF57301.
XX Novel FCR3, varCSA protein, useful for modulating parasitized red blood

PT cell binding, sequestration and onset of maternal malaria -
 PS Claim 12; Page 63-71; 78pp; English.
 CC The invention relates to a P. falciparum FCRI3 varCSA protein, that is
 CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
 CC corresponding P.falciparum erythrocyte membrane protein 1 (PFEMP1)
 CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
 CC protein and the encoding gene are useful for treating and preventing
 CC maternal malaria in a patient identified at a risk for contracting
 CC maternal malaria or in a patient afflicted with maternal malaria. The
 CC present sequence represents the P. falciparum FCRI3 varCSA protein.
 SO Sequence 3542 AA;
 Query Match 100.0%; Score 1918; DB 22; Length 3542;
 Best Local Similarity 100.0%; Pred. No. 3e-138;
 Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 YEKEISYLSNDKMFVNINSEYKQFEKLETOYATNDFTLNLNEGKYGKGLPGEK 60
 403 YEKEISYLSNDKMFVNINSEYKQFEKLETOYATNDFTLNLNEGKYGKGLPGEK 462
 OY 61 DIFETNSADDKGIFYSREYCOVCPDGVKCDGKIKYTHKSDNDEYRNDEYKPPWGVKPT 120
 463 DIFETNSADDKGIFYSREYCOVCPDGVKCDGKIKYTHKSDNDEYRNDEYKPPWGVKPT 522
 OY 121 NITVALSGNEQDITQKLEFNCSSYNYKDKNNQKWEYKDEINRCKLEQTEINNDN 180
 523 NITVALSGNEQDITQKLEFNCSSYNYKDKNNQKWEYKDEINRCKLEQTEINNDN 582
 OY 181 PKTISHNFELWYTLADDTIKMNDKLTQCIINNTTHCIDCNRCCLCFDRWVQKEE 240
 583 PKTISHNFELWYTLADDTIKMNDKLTQCIINNTTHCIDCNRCCLCFDRWVQKEE 642
 OY 241 WNSIKKLFKKKNIQOOSYSNINLFEYGFYFVMDKLDDEAKWKEIMENIKRKNFEFSN 300
 643 WNSIKKLFKKKNIQOOSYSNINLFEYGFYFVMDKLDDEAKWKEIMENIKRKNFEFSN 702
 OY 301 LENNBDYLENALELLDHLKETATTCCKDNTEACETSHNATNP 345
 703 LENNBDYLENALELLDHLKETATTCCKDNTEACETSHNATNP 747
 DB
 RESULT 2
 AAM93944
 ID AAM93944 standard; protein: 2228 AA.
 XX AAM93944;
 30-JUN-1999 (first entry)
 DE P. falciparum PfEMP1 protein.
 XX Erythrocyte membrane protein; EMP; PfEMP1; malaria; antioclusional;
 KW glycosamino-glycan-like moiety; antiaggregational; antimalarial;
 KW antigen receptor; infected erythrocyte; rosette formation; blood cell;
 KW capillary occlusion; cerebral malaria; treatment; vaccine; detection;
 KW medicament; parasite; diagnosis; drug screening.
 OS Plasmodium falciparum.
 XX WO9915557-A1.
 XX 01-APR-1999.
 XX 18-SEP-1998; 98WO-SE01675.
 XX 19-SEP-1997; 97SE-0003386.
 PA (KARO-) KAROLINSKA INNOVATIONS AB.
 XX Barragan A, Carlson J, Fernandez V, Oljun C, Wahlgren M;

XX MPI: 1999-254692/21.
 DR New isolated malaria polypeptides
 PS Claim 4; Page 67-74; 80pp; English.
 CC This invention describes a novel Plasmodium falciparum erythrocyte
 CC membrane protein (EMP), PfEMP1, which is capable of binding to a
 CC carbohydrate which exhibits at least one negatively charged
 CC glycosamino-glycan (GAG)-like moiety and has antiaggregational,
 CC antioclusional and antimalarial activity. The carbohydrates of the
 CC invention are capable of acting as receptors for malaria antigens
 CC present on the surfaces of malaria infected erythrocytes, by binding
 CC to these antigens the carbohydrates prevent rosette formation by the
 CC blood cells, this prevents occlusion of capillaries as is seen in
 CC cerebral malaria caused by Plasmodium falciparum. The products of the
 CC invention can be used to treat malaria or to vaccinate against it, or
 CC used to design a model to identify compounds that bind to PfEMP1. The
 CC carbohydrates, polypeptides and antibodies of the invention can be used
 CC as a medicament for dissolving the rosettes formed by erythrocytes
 CC infected by a malaria parasite. The products can also be used for
 CC detection, diagnosis and drug screening.
 SO Sequence 2228 AA;
 Query Match 23.3%; Score 446; DB 20; Length 2228;
 Best Local Similarity 31.4%; Pred. No. 3.7e-25;
 Matches 131; Conservative 53; Mismatches 133; Indels 100; Gaps 20;
 OY 1 YEKEISYLSNDKMFVNIN--SEYKQFEKLETOYATNDFTLNLNEGKYGK----- 53
 415 YEKEISDGGRRKRAVGTTKYEKESFEKTKMNDGIGYDAFLGLLNKACKDITDG 474
 OY 54 -----GGLPGEKDTFTNSADDKGIFYSREYCOVCPDGVKCDGKIKYTH 97
 475 GKINKEVNSGGGVGGSGGTGSGASGIND--ENKGTFRSEYCOVCPDGVK-----H 526
 OY 98 KSDNDEYRNDEYKPPG--VKPTN--ITVALSGNEQD---ITQKLEFNC--NSS- 145
 527 KGNQWEM-KTKYKMKRMSKIKYKPIKNGAMVLLSLAKYVKMTLKNWKEFLTONSSD 585
 OY 146 -----TNYKDKNNQK-----WECYKDEINRCKLEQTEINND----- 179
 586 GSVGSVYVTGASGNSSEKELYDEMKC-YKINEYQXNVQGEVEDEDELKAGAGLITLP 644
 OY 180 NPK-----IISFHNFEFLWYTLADDTIKMNDK-LKTCINN-TTTHCI 220
 645 NPKNKEVSEAKSONNHADLOKTFHDEFFYVVAHMLKDSIMWTRKRKSCISDGTWKR 704
 OY 221 DECNRNCLCFDRWVQKQKEEWSIKLFTKKKNIQOSY-SNINLFEYGF-----V 273
 705 NGCNKRCDCFEKWKQKETEKKPIKDHFKTOEGIPBGIYTTTLELILKLOFLEDEENT 764
 OY 274 MDKLDDEAKWKEIMENIKRKNFEFSNLE--NNRDYLENALELLDHLKETATTCCKD 328
 765 ENSIDAEAEELKHQIKILKLENE--NMLAVNAGTEQKTLMDKLNLNELDARCKMD 820
 DB
 RESULT 3
 AAB18352
 ID AAB18352 standard; protein: 2197 AA.
 XX AAB18352;
 XX 07-NOV-2000 (first entry)
 DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:210.
 XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoacide; infection; insecticide.
 OS Plasmodium falciparum.

[illegible]

DB	740	---	LMDSADVLELLELEQLFQDTRKDGVDKELKGIKELLDEKKKKQAEAVVYVA	796
QY	302	ENNRDYLENAIELLDHLEKETAATICKDNNTNCE	336	
DB	797	DNOK---KTTIDKILQHGGDANNCLKTH-KEKCE	827	
		RESULT 4		
ID	AAB6344	standard; protein; 445 AA.		
AC	AAB6344;			
XX	05-APR-2001	(first entry)		
DE		Malarial parasite derived protein pf-EMP-1 fragment CIDR/A4.		
XX		Malarial parasite; pf-EMP-1; vaccine; malaria; autoimmune disease;		
XX		dendritic cell maturation; immune tolerance.		
OS		Unidentified.		
XX		WO200102005-A2.		
PN		11-JAN-2001.		
PD		30-JUN-2000; 2000WO-GB02546.		
PF		30-JUN-1999; 99GB-0015311.		
PR		(ISIS-) ISIS INNOVATION LTD.		
PA		Roberts DJ, Urban BC;		
PI		WPI; 2001-138065/14.		
DR				
XX		Treating mammalian dendritic cells in vitro to induce immune tolerance		
PT		comprising exposing the cells to an agonist of the cell surface		
PT		receptors CD36 and/or CD51, useful for treatment of autoimmune		
PT		disorders -		
XX		Claim 47; Fig 2; 148pp; English.		
PS				
CC		The present invention describes methods of inducing immune tolerance in		
CC		dendritic cells by exposing the cells to an agonist of the cell surface		
CC		receptors CD36 or CD51, and of identifying molecules capable of		
CC		preventing the adhesion of red blood cells infected with a malarial		
CC		parasite to dendritic cells. The latter method involves the exposure of		
CC		CD36 to the test molecule and parasitised human red blood cells, and		
CC		determining the level of adherence of the blood cells to the dendritic		
CC		cells. A molecule which causes a reduction in the adherence is useful in		
CC		the treatment of malaria. The first method is useful in the treatment of		
CC		autoimmune diseases.		
SQ	Sequence	445 AA;		
QY	Query Match	22.5%; Score 431; DB 22; Length 445;		
DB	Best Local Similarity	28.1%; Pred. No. 6.3e-25;		
	Matches 109; Conservative 71; Mismatches 134; Indels 74; Gaps 13			
QY	11	NDNRKV---NNINEEYKQFEYKLEKQYATNDPFLNLNLNGKYYCKG---GLDGEKDI 62		
DB	7	NDDTITTEYGINNMRYKRDYFKHLEE-KYKTVDAFLNLNLNKEKEKNHPEVGEKKKI 65		
QY	63	TFTNSADDKGIFRSEYQVCPDGCYKCDGIKYTHKSDRDRVNN-----EDYKPPW 115		
DB	66	DENDNIE--TFSHTEYRPPCPWCGIE-----EDKDGKMRINDHSACKKEELYTPKE 115		
QY	116	GKPKINIVLYSGNNGODITQKLEFNCNSSTYYKDKNN-----QKMECYTKDE- 163		
DB	116	NAKYRIANVLTSGEGHEDIARLKEFCYTYONGGGGSDDCGNSDSISLCEPQCYQPDQL 175		

QY 164 -----NINRCKLEQNTPEINNDPKIISFHNFFELWVYLLRDTIKW 204
 Db 176 EKVGGEVDKLGAGGLCEFEKMKGEKVK-----KQTFNPFNFWAHVLKDSIDW 229
 QY 205 NDKLKTGIN-NTTHCIDECDNRNCLCFDRWVKQKEEEMNSIKLFTKKK-100SYISNI 262
 Db 230 RTQTLTKLSEKLLKCKEKGKSNCECKRKWKKEKEMIKVKQDFNKTDFLEKRVHVL 289
 QY 263 NNLEGEYFFKVMKLDKDEAKWKELMENIKRKKNEFSNLNRRDYLENALTELLDHLKET 322
 Db 290 ETLLENYFFENIKAYGDLKSLIOEMKKMIKENKQNKRTD-----EDALDYLFDEKEE 345
 QY 323 ATIC-----KDNNTNEACETSHNATNP 345
 Db 346 AEDCLDIHEDDDDDCEVELEKIPNP 373

RESULT 5
 AAR70236
 ID AAR70236 standard; Protein; 2703 AA.

AAR70236;
 DT 22-SEP-1995 (first entry)
 XX
 DE P. falciparum Proj3.
 XX
 KW Erythrocyte binding ligand; Proj3; binding domain; malaria; therapy;
 XX vaccine.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO9507353-A.
 PD 16-MAR-1995.
 XX
 PF 07-SEP-1994; 94WO-US10230.
 XX
 PR 10-SEP-1993; 93US-0119677.
 XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Chlunis C, Miller LH, Peterson DS, Slim KL, Su X;
 XX Wellens TE;
 DR WPI: 1995-123427/16.
 DR N-PSDB; AAO83529.
 XX

New erythrocyte binding domain polypeptide(s) - isolated from
 Plasmodium binding proteins, used in diagnosis, treatment and
 prevention of malaria
 XX
 PS Disclosure: Page 61-65; 81pp; English.
 XX

CC Erythrocyte binding ligand (EBL) family genes were cloned from
 CC P. falciparum chromosome 7 subsegment libraries constructed during
 CC genetic studies of the chloroquine resistance locus. The 4 genes,
 CC EBL-e1 (AAO83526), E3la (AAO83527), EBL-e2 (AAO83528) and Proj3
 CC (AAO83529), encode the proteins given in AAR70233-36, respectively. The
 CC binding domains of such proteins can be expressed e.g. in E. coli,
 CC yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
 CC cells, and provide protection against P. falciparum.
 XX

XX Sequence 2703 AA;

Query Match 20.6%; Score 394.5; DB 16; Length 2703;
 Best Local Similarity 26.2%; Pred. No. 4.4e-21;
 Matches 122; Conservative 76; Mismatches 135; Indels 133; Gaps 20;

QY 1 YEKEIOSY-----LSNNKFPVNNINSE-YKQFEKLEKETOYATNDFMLINLNEK 50
 Db 389 YDEIKKREKNAAGSGSRKRAAGGTTTNNYGEKFEYDELNKSEYRTVDFLEKLSNEE 448

QY 51 YC-----KGLPEKDDITFTNSA-----DDKGIFFSERCQVCPDGKCDGIKTYH 97
 Db 449 ICTKVADEEGGTTIDFKNVNSDSTSGASGTVNESQGFYSKYCOQPCPYCGVK-----KVNN 504
 QY 98 -KSDNDRERVNN-----EDYKPPWGVKPPNITVLVSGNDOGDITQKLENFCN-----143
 Db 505 GGSNSNMEENKNNCKSKGIYEPKPKREGTTITLLSGGHDDIEKLLKFKDEKNGDTI 564
 QY 144 -----SSTNTKDKNN-----QKWCYKNDENINRCKLEQNTPEINNDPK-----11-- 184
 Db 565 NSGSGTGGSGGNSGSGRQELYEEMKC-YKGEDEVKVGHEDEDEYENKNNAGLCILIKN 623
 QY 185 -----SFHNFELWVYLLRDTIKAKDKLTKCI-NTTTHC-ID 221
 Db 624 QKKNKEGGNTSEKPEDELOKTFNPFYVVAHMLKDSITHMKILQRCLONGNRICKGN 683
 QY 222 EGNRNCLEDRWVKQKEEEMNSIKLFTKKNIQ-----QSYSNINLFE 267
 Db 684 KCNNDCECFKRWLTQKKDKMGKIYQHF-KTONIKRGSGSDNTAELLIPDHDYVLOYNLOE 742
 QY 268 GYF-----FKWADKLDKDEAKWKELMENIKRKKNEFSNLNRRD-----YLEN 310
 Db 743 EFLKGDSEDASEKSENSIDAEAA-----EELKHLREITSEDNNOEASVGGVTEQKN 796
 QY 311 ATELLDHLKETATIC-----KDNNTNEACETSHNATNP 345
 Db 797 IMDKLLNFKEDADLCLEIHEDEBEKEKGDGNECIEGENRYPN 842

RESULT 6
 AAW22482
 ID AAW22482 standard; Protein; 2710 AA.
 AC AAW22482;
 XX

DT 07-OCT-1997 (first entry)
 XX
 DE Plasmodium Proj3.
 XX

KW DBL gene family; SAMP: stalic acid binding protein; vaccine; therapy;
 KW DABP; mezoizole; malaria; var-1; var-2; var-3; var-7; immune response;
 KW Plasmodium.
 XX

OS Plasmodium falciparum.
 XX

PN WO9640766-A2.
 XX

PD 19-DEC-1996.
 XX

PF 07-JUN-1996; 96WO-US09508.
 XX

PR 07-JUN-1995; 95US-0487826.
 XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX

PI Chlunis C, Miller LH, Peterson DS, Slim KL, Su X;
 XX Wellens TE;
 DR WPI: 1997-052231/05.
 DR N-PSDB; AAT72897.
 XX

PT New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and stalic acid
 PT binding proteins
 XX

PS Disclosure: Page 50-56; 96pp; English.
 XX

CC This sequence represents Proj3 of Plasmodium. Proj3 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and stalic acid binding protein
 CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 CC var family of genes modulate cytoadherence and antigenic variation of

30-NOV-1999.

10 MAMZ24/3 STANBARDU; FLOREIN; 3000 MW.

AAW22475;
 12-SEP-1997 (first entry)
 Plasmodium var-7.
 DBL gene family; SABP; stalic acid binding protein; vaccine; therapy;
 Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 Plasmodium.
 OS Plasmodium vivax.
 OS Plasmodium falciparum.
 W09640766-A2.
 19-DEC-1996.
 07-JUN-1996; 96WO-US09508.
 07-JUN-1995; 95US-0487826.
 (USSH) US DEPT HEALTH & HUMAN SERVICES.
 Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 Wellens TE;
 WPI; 1997-052331/05.
 N-PSDB; AAT72882.
 New malaria vaccines - contains cysteine-rich DBL family protein
 binding domains homologous domains of the Duffy and stalic acid
 binding proteins
 Claim 8; Page 61-67; 96pp; English.
 This sequence represents var-7 of Plasmodium. Var-7 belongs to
 the Duffy binding like (DBL) family of genes which have homology to
 Duffy antigen binding protein (DABP) and stalic acid binding protein
 (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 var family of genes modulate cytoadherence and antigenic variation of
 Plasmodium infected erythrocytes. SABP and the duffy antigen binding
 protein (DABP) are soluble proteins that appear in the culture
 supernatant after infected erythrocytes release merozoites. DABP and
 SABP mediate the binding of merozoites and schizonts to the erythrocyte
 surface. These proteins are necessary for erythrocyte invasion by the
 parasite. This sequence can be used in the compositions of the
 invention. The compositions are for the treatment and prevention of
 malaria, and comprise either a nucleotide sequence or encoded polypeptide
 of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
 family of genes having homology with conserved regions of DABP and SABP.
 The compositions are used for the treatment and prevention of malaria.
 They are also used in the preparation of vaccines for inducing a
 protective immune response in a mammal to Plasmodium merozoites
 (especially Plasmodium falciparum or Plasmodium vivax).
 Sequence 3060 AA;
 Query Match 20.6%; Score 394.5; DB 18; Length 3060;
 Best Local Similarity 26.2%; Pred. No. 5.2e-21;
 Matches 122; Conservative 76; Mismatches 135; Indels 133; Gaps 20;
 1 YEKETQST-----LSNDNKFVNINSE-YKQFEKELKNOYANDNFILNLSGK 50
 Db 394 YDEIKKTKENASGSRKRAAGCTTTNYYGKEKFDLKNKSTRYVDKFLKLSMEE 453
 QY 51 YC-----KGLPEKEDITFTNSA-----DKGIFYRSEYCOVPCDGCIGIKYTH 97
 Db 454 ICTFKYDEGGTIDFKNVNSDSTGASCTNVSQGTFRSKYCCPCPCGK-----KYN 509
 QY 98 -KSNDRNRVNN-----EDYKPPMGVAPNINIVLYSGNEGDITQKLENFCN----- 143

510 GSSNMEMEKNKCKCKGKIYEPKPEKGTITITLKSQGHDIIEKLNKFEDEKNGDTI 569
 QY 144 -----SSTNFKDKN-----QWECYRDENINRCKLEONTLEINNDPK-----II-- 184
 Db 570 NSGSGTGGSGGNSGRQELYEEMK-YKGEDEVKVGHEDEDEDEYENKKNAGCICLKN 628
 QY 185 -----SFHNFELAVYLLRDTIKWMDKLTCTI-NNTTNHC-ID 221
 Db 629 QKNKKEGGNTSEKPEDEIQTFNPFYYVAHMLKDSIHMKKIQRCONRRIKCGN 688
 QY 222 EGNRNCICEDRWYKQKEEWSNISKLTFTKKNIO-----QSYSNINLNEE 267
 Db 689 KGNNDCECFKRMITQKKDMGKIYQHF-KTQNIKGGSDNTAELIPHDYVLYOINLOE 747
 QY 268 GYF-----FKYWDKDKDEAKKKEIMENIKRKNFESMLENND-----YLEN 310
 Db 748 EFLKGDSEDAESEKSENSIDAEEA-----EELKHLREIESEDDNNQASVGGVTEQKN 801
 QY 311 ATEILLHLKETATTC-----KDNNTNEACENSHNATTP 345
 Db 802 IMDKILNTEKDEADLCLEIHEDEBEKEKGDGNECIEGSENRYP 847
 RESULT 9
 AAY77905
 ID AAY77905 standard; Protein; 3060 AA.
 XX AAY77905;
 DT 13-JUN-2000 (first entry)
 XX Plasmodium var-7 polypeptide.
 DE DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
 KW DABP; Stalic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
 KM protozoacide; var-7.
 XX Plasmodium sp.
 OS US5993827-A.
 XX 30-NOV-1999.
 PE 07-JUN-1995; 95US-0487826.
 XX 10-SEP-1993; 93US-0119677.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;
 WPI; 2000-194198/17.
 N-PSDB; AA298287.
 Isolated protein binding domains from Plasmodium vivax and Plasmodium
 falciparum erythrocyte binding proteins useful for vaccinating against
 malaria
 Disclosure; Columns 109-124; 93pp; English.
 The invention relates to ebl-1 polypeptides that are encoded by the DBL
 (Duffy-binding like) gene family. The ebl-1 proteins are substantially
 identical to the Duffy Antigen Binding Protein (DABP) and Stalic Acid
 Binding Protein (SABP), which are soluble proteins that appear in the
 culture supernatant after erythrocytes infected with malaria release
 merozoites. Immunochemical studies indicate that DABP and SABP are the
 respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
 and stalic acid receptors on erythrocytes. The ebl-1 polypeptides may be
 used to vaccinate against malaria, especially caused by P. falciparum.
 Immunization with the polypeptide provides effective protection against
 malaria. The present sequence represents the var-7 polypeptide.
 Sequence 3060 AA;

Query Match 20.6%; Score 394.5; DB 21; Length 3060;
 Best Local Similarity 26.2%; Pred. No. 5.2e-21;
 Matches 122; Conservative 76; Mismatches 135; Indels 133; Gaps 20;

1 YEKEIOSY-----LSNDKFNANNINSE-YKQEFYEKLTQYATNDFTLNECK 50
 394 YDEIKKYENGASGSRORADAGITTYNDGEKFFYDELNKSEYRTYDKELEKLSNEE 453
 51 YC-----KGLPGKEDITFTNSA-----DDKGFYRSEYCOVCPDGVKCDGIRK 97
 454 ICTKVADESGTIDFNKVNSDSTSGASGTNVEGQFFYRSKYQPCPGCVK-----KVN 509
 98 -KSDNREKVN-----EDYKPPGVKPTNTVLYSGNEQGITQKLEFCN-----143
 510 GGSNNMEKNNCKCKSGKLYEPKPEGTITILKSGKHDIIEKLNKFCDEKNDITL 569
 144 -----SSTNFKDKNN-----QKMECYKDENNRCKLEQNTLEINDNPK-----114
 570 NSGSGTGGSGGSGRQELIYEMKC-YKGEDEVKVGHDDEDEDEVNNAAGLCILKN 628
 185 -----SFHNFPELWVLYLRDTIKMNDKLTQCI-NNTTTC-ID 221
 629 QKKNKREGNTSEKPEDELQKTFNPFYVVAHMLKDSIMKKKLRCLONGRIRKCGN 688
 222 ECKRNLCDRWVYKQKEEEMNSIKLFTYKKKNIQ-----QSYYNINNLFE 267
 689 KCNNDCEKRWITOKKDEGKIYQHF-KTONIKRGSGSDNTAELLPEFDHYLYQNLQ 747
 268 GYF-----FKVMDKLDKDEAKWKELMENIKRKKNEHSNLENNRD-----YLEN 310
 748 EFKGSGEDASEKSEKSENDAEBA-----EELKHLREITTESDNNDEASVGGVTEQKN 801
 311 ALELLDLHKEPATIC-----KDNNTNEACETSHNATTNP 345
 802 IMDKLTNKEADLCLTEIHEDEEKEKGDGNECIEEGENFRYPN 847

RESULT 10
 AAR70235
 ID AAR70235 standard; Protein; 700 AA.
 XX AAR70235;
 AC
 XX
 DT 22-SEP-1995 (first entry)
 XX
 DE P. falciparum EBL-e2.
 Erythrocyte binding ligand; EBL-e2; binding domain; malaria; therapy;
 vaccine.
 XX Plasmodium falciparum.
 OS
 XX
 PN W09507353-A.
 PD
 XX 16-MAR-1995.
 XX
 PF 07-SEP-1994; 94WO-US10230.
 XX
 PR 10-SEP-1993; 93US-0119677.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 XX
 DR WPI: 1995-123427/16.
 DR N-PSDB; AAO83528.
 XX
 XX New erythrocyte binding domain polypeptide(s) - Isolated from
 PT Plasmodium binding proteins, used in diagnosis, treatment and
 PT prevention of malaria
 XX

PS Disclosure; Page 56-57; 81pp; English.
 XX
 CC Erythrocyte binding ligand (EBL) family genes were cloned from
 CC P. falciparum chromosome 7 subsegment libraries constructed during
 CC genetic studies of the chloroquine resistance locus. The 4 genes,
 CC EBL-e1 (AA083526), E3ia (AA083527), EBL-e2 (AA083528), and Pro13
 CC (AA083529), encode the proteins given in AAR70233-36, respectively. The
 CC binding domains of such proteins can be expressed e.g. in E. coli,
 CC yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
 CC cells, and provide protection against P. falciparum.
 XX

Sequence 700 AA:

Query Match 19.8%; Score 380; DB 16; Length 700;
 Best Local Similarity 28.7%; Pred. No. 9.7e-21;
 Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

1 YEKEI-----QSYLSNDKFNANNINSEYKQEFYEKLTQYATNDFTL 43
 285 YETELISGGSGKSPKRTKRAARSSSSDDN-----GYESKFKKLKEGVQYDVDFKL 336
 44 NLNNEGYCKG-----GLPGKEDITFTNSADDKGFYRSEYCOVCPDGVKCDGIRK 99
 337 KIINKESICQKQOVGNKADNDFTNEKYK-TFSKTEICPCPCGCLKGGPPKVKG 395
 100 DNDREKVNEDYKPPMGVCPNTVLYSGNEQGITQKLEFCNCSSTNYKDKNNQMECY 159
 396 DKTCGSAKRTYDPK---NTDIPVLYPDKSQQNLKKNKFCBKAP-GGQIKKKQCY 451
 160 YKD-----ENINCKLEQNTLEINDNKKIISFHNFEFLWVLYLRDTIKMNDKLTQ 214
 452 YDEHRPSSKNNNCKVEGTWKTQKQYKSYNVEFDWHDMLHDSVEKTELSKINN 511
 215 TTT---HCIDECNRNLCDRWVYKQKEEEMNSIKLFTYKKKNIQSYYNINNLFE 270
 512 NTNGNTCRNNKCKTDCGCKQKWEKKQEMAKKHDFGQTDVQO-----KGLIYFSP 566
 271 FKVMDKLDKDEAKWKELMENIKRKKNEHSNLENNRDYLE-----NALELL 316
 567 YGVLDIV---LKGGNLQNIKDVHGDTDIDIKHKKLLDEDAVAVVLGGKNDTTIDKL 622
 317 DILKETFATICKDNTNTFACE 336
 623 QHEKEQAEOCKOR--OECE 640

RESULT 11
 AAW22481
 ID AAW22481 standard; Protein; 700 AA.
 XX AAW22481;
 AC
 XX
 DT 07-OCT-1997 (first entry).
 XX
 DE Plasmodium ebl-2.
 XX
 KW DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
 KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 XX Plasmodium.
 XX Plasmodium falciparum.
 OS
 XX
 PN W09640766-A2.
 PD
 XX 19-DEC-1996.
 XX
 PF 07-JUN-1996; 96WO-US09508.
 XX
 PR 07-JUN-1995; 95US-0487826.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX

PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 DR WPI; 1997-052231/05.
 DR N-PSDB; AAT72896.
 XX
 PT New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and sialic acid
 PT binding proteins
 XX
 PS Disclosure; Page 46-48; 96pp; English.

This sequence represents ebl-2 of Plasmodium. Ebl-2 belongs to
 the Duffy binding like (DBL) family of genes which have homology to the
 Duffy antigen binding protein (DABP) and sialic acid binding protein
 (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 var family of genes modulate cytoadherence and antigenic variation of
 Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 protein (DABP) are soluble proteins that appear in the culture
 supernatant after infected erythrocytes release merozoites. DABP and
 SABP mediate the binding of merozoites and schizonts to the erythrocyte
 surface. These proteins are necessary for erythrocyte invasion by the
 parasite. This sequence can be used in the compositions of the
 invention. The compositions are for the treatment and prevention of
 malaria, and comprise either a nucleotide sequence or encoded polypeptide
 of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
 family of genes having homology with conserved regions of DABP and SABP.
 The compositions are used for the treatment and prevention of malaria.
 They are also used in the preparation of vaccines for inducing a
 protective immune response in a mammal to Plasmodium merozoites
 (especially Plasmodium falciparum or Plasmodium vivax).

Sequence 700 AA:

Query Match 19.8%; Score 380; DB 18; Length 700;
 Best Local Similarity 28.7%; Pred. No. 9.7e-21;
 Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

1 YKEI-----QSLNDNKRNVNINSERYKQFEKLEKETOYATNDTFL 43
 285 YETELISGGSGSPKRTKRAARSSSSDDN-----GESKRYKKLEKGYQDVDFKL 336
 44 NLNMGKCYCG---GLPEKEDITFTNSADDKGIFYSRQYQVPCGVCVKGCIGITYTHKS 99
 337 KILNKEGICQKQPOYGENEKADNDVFTNEKYK-TFSRTLEICPCWCGLEKGGPPWKYKG 395
 100 DNDREVRNNEDEKPPWGVKPFNTIVLYSGNEQGITQKLEFNCSSSTNYKDKNNQWECY 159
 396 DKTGSAKRTYDPR---NITDIPVLYPDKSQNIIKKYKNCCEKGP-GGGQIKKWQCY 451
 160 YKD-----ENINRCKLEQNTNNDNPKIISPHNEFELWVYLLDPTIKMNDKLTCTINN 214
 452 YDEHRPSSKNNNNNCVEGTWDFKQKQYKSYNVEFMDVHDLHDSVEKTELSKCIINN 511
 215 TTT---HCIDECNRNCLCFDRWVYKQKEEEMNSIKKLTFTKKNIQOQSYISNINFEGYF 270
 512 NTNGNTCRNNKCKTDGCGCFQKWEKKQOEWMAIKQHPFGKQTDIVQO-----KGLIVFSP 566
 271 FKVMNKLNDKDEAKKELMENIKRKKNEFSNLNNRDYLE-----NAIETLL 316
 567 YGVLDLY---LKGNNLQNIKQVHGDPTDIKIKILDEBDAVAVVLGKDKNTTIDKLL 622
 317 DILKETATICKDNNTNENACE 336
 623 QHEKQAEQCKQK--QECE 640

RESULT 12

AAAT77903
 ID AAT77903 standard; Protein; 700 AA.
 XX
 AC AAT77903;
 XX

DT 13-JUN-2000 (first entry)

XX P. falciparum ebl-2 polypeptide.
 DE
 XX DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
 KW DABP; Sialic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
 KM protozoacide.
 XX

OS Plasmodium falciparum.

XX US593827-A.

XX 30-NOV-1999.

XX 07-JUN-1995; 95US-0487826.

XX 10-SEP-1993; 93US-0119677.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;

DR WPI; 2000-194198/17.

DR N-PSDB; AAZ98285.

PT Isolated protein binding domains from Plasmodium vivax and Plasmodium
 PT falciparum erythrocyte binding proteins useful for vaccinating against
 PT malaria -

XX Disclosure; Columns 67-72; 93pp; English.

The invention relates to ebl-1 polypeptides that are encoded by the DBL
 (Duffy-binding like) gene family. The ebl-1 proteins are substantially
 identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
 Binding Protein (SABP), which are soluble proteins that appear in the
 culture supernatant after erythrocytes infected with malaria release
 merozoites. Immunochemical studies indicate that DABP and SABP are the
 respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
 and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
 used to vaccinate against malaria, especially caused by P. falciparum.
 Immunization with the polypeptide provides effective protection against
 malaria. The present sequence represents the ebl-2 polypeptide.

Sequence 700 AA:

Query Match 19.8%; Score 380; DB 21; Length 700;
 Best Local Similarity 28.7%; Pred. No. 9.7e-21;
 Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

1 YKEI-----QSLNDNKRNVNINSERYKQFEKLEKETOYATNDTFL 43
 285 YETELISGGSGSPKRTKRAARSSSSDDN-----GESKRYKKLEKGYQDVDFKL 336
 44 NLNMGKCYCG---GLPEKEDITFTNSADDKGIFYSRQYQVPCGVCVKGCIGITYTHKS 99
 337 KILNKEGICQKQPOYGENEKADNDVFTNEKYK-TFSRTLEICPCWCGLEKGGPPWKYKG 395
 100 DNDREVRNNEDEKPPWGVKPFNTIVLYSGNEQGITQKLEFNCSSSTNYKDKNNQWECY 159
 396 DKTGSAKRTYDPR---NITDIPVLYPDKSQNIIKKYKNCCEKGP-GGGQIKKWQCY 451
 160 YKD-----ENINRCKLEQNTNNDNPKIISPHNEFELWVYLLDPTIKMNDKLTCTINN 214
 452 YDEHRPSSKNNNNNCVEGTWDFKQKQYKSYNVEFMDVHDLHDSVEKTELSKCIINN 511
 215 TTT---HCIDECNRNCLCFDRWVYKQKEEEMNSIKKLTFTKKNIQOQSYISNINFEGYF 270
 512 NTNGNTCRNNKCKTDGCGCFQKWEKKQOEWMAIKQHPFGKQTDIVQO-----KGLIVFSP 566
 271 FKVMNKLNDKDEAKKELMENIKRKKNEFSNLNNRDYLE-----NAIETLL 316
 567 YGVLDLY---LKGNNLQNIKQVHGDPTDIKIKILDEBDAVAVVLGKDKNTTIDKLL 622

CC Immunization with the polypeptide provides effective protection against
CC malaria. The present sequence represents the var-1 polypeptide.

SO Sequence 2182 AA;

Query Match 19.8%; Score 380; DB 21; Length 2182;
Best Local Similarity 28.7%; Pred. No. 4.3e-20;
Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

```
QY 1 YEKEL-----QSLSDNKKFVNINSEYKQFEKLETOYATNDFL 43
  |||
DB 413 YETELSGGSGSKSPKRTKRAANSSSDN-----GYESKFTYKLEKVGTDYDFL 464
  |||
QY 44 NLNKGKXCKG---GLPGEKIDFTNSADKGFYRSEYCCVPCDCCGKCDGKITYHKS 99
  |||
DB 465 KINNEGICQKQPOYGNENKADVFTNEKYK-FTSRTEICPCPCWCGLEKGGPPMKVKG 523
  |||
QY 100 DNDREVRNNEYKPPWGVKFTNITVLYSGNEGDTQKLENCSNTAKKNNOKMEY 159
  |||
DB 524 DKTCGSAKTKYDPK---NITDIPVLPKRSQONTLKTKKNCCKGAP-GGQIKRMOCY 579
  |||
DB 160 YKD-----ENIRCKLEQNTLEINNDNPKIISFNFEFLVYTLRDTIKMNDKLTGINN 214
  |||
DB 580 YDEHPSSKNNNVCYEGTWDKFTQKQYKSTNVEFWMDVHDLDSYEMKTELKCIIN 639
  |||
QY 215 TTT-----HCIDECNRNCLCFDRWVKQKEEWSIKKLFKKKNIOQSYYSNINLEFGEY 270
  |||
DB 640 NTNGWTCNRNNCKTDCGCFQKWEKQOEWMAIKDHFGKQDIYQ-----KGLVESP 694
  |||
QY 271 FRVMDKLDKDEAKKLENIKRNKNSLENNDYLE-----NAIEILL 316
  |||
DB 695 YGVLDLV-----LKGNLQNIKIVHGDITDIKIKKLDDEDAVAVVLGKNDTITDKLL 750
  |||
QY 317 DLKLEFATICKDNTNEACE 336
  |||
DB 751 QHEKEQAECKOK--QECE 768
  |||
```

RESULT 15
ID AAM00385.
AAM00385 standard; Protein; 1726 AA.

AC AAM00385;
XX
DT 21-FEB-1997 (first entry)
XX
DE Truncated Plasmodium falciparum erythrocyte membrane protein.
XX
Plasmodium falciparum; erythrocyte membrane protein; malaria;
detection; identification; treatment; prevention; parasite.

OS Plasmodium falciparum MC type.
XX
XX
FH Key Location/Qualifiers
FH Domain 62..394
FH /label= Duffy binding ligand domain 1
FH Region 607..648
FH /note= "Cysteine rich motif"
FH Domain 839..1282
FH /label= Duffy binding ligand domain 2
FH Region 1488..1523
FH /note= "Cysteine rich motif"

MO96373736-A1.
XX
XX
PD 31-OCT-1996.
XX
XX
PF 26-APR-1996; 96WO-US05798.
XX
XX
PR 27-APR-1995; 95US-0430908.
XX
XX
PA (AFFY-) AFFYMAX TECHNOLOGIES NV.
XX

PI Baruch DI, Howard RJ, Pasloske BL;
XX
XX WPI: 1996-497376/49.
DR N-PSDB; AAT41853.
XX

PT New Plasmodium falciparum erythrocyte membrane proteins - used to
PT develop products for the diagnosis, treatment or prevention of
PT malaria parasite infections
XX
XX
PS Claim 1; Figure 12; 149pp; English.

CC A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte
CC membrane protein 1 (PfEMP1) or active fragments or analogues of that
CC protein can be used in the treatment or prevention of symptoms of a
CC malaria parasite infection. The polypeptides can inhibit, block or
CC reverse the sequestration of erythrocytes in patients suffering from
CC malaria. Nucleic acids derived from the PfEMP1 gene can be used as
CC probes and primers to identify a Plasmodium falciparum parasite, the
CC primers used to generate characteristic amplification patterns from
CC different P. falciparum strains. Antibodies specifically
CC immunoreactive with the PfEMP1 polypeptide or its fragments may be
CC used in diagnosis of malaria infection. This is a truncated PfEMP1
CC protein of the MC type of Plasmodium falciparum. The full length
CC PfEMP1 protein is given in AAM00384.
XX
XX

SO Sequence 1726 AA;

Query Match 18.6%; Score 357.5; DB 17; Length 1726;
Best Local Similarity 27.4%; Pred. No. 1.7e-18;
Matches 104; Conservative 64; Mismatches 137; Indels 75; Gaps 16;

```
QY 1 YEKELQ---SYLSDNKKFVNINSEYKQFEKLETOYATNDFLNLNKGKXCK--G 54
  |||
DB 394 YDEELNKTHGTTITGNGKINNL---YVHFYKILKK-YTPVDSKQKLNDEALCKRP 449
  |||
QY 55 GLPGEKIDFTNSADKGFYRSEYCCVPCDCCGKCD---GKITYHKS-DNDREVRNNE 109
  |||
DB 450 NVGNEKASTVDENNEVNTFTSHTYCEACPCWCGAKENGGKAKKESCAKKEKREIFNK 509
  |||
QY 110 DYKPPGVKFTNITVLYSGNEGDTQKLENCSNTAKKNNOKMECYKDEINRCK 169
  |||
DB 510 E-----NSTDIKITLPKGRSKTLEKTKTCKQKGRK---NDIWKCHYDNG----- 554
  |||
QY 170 LEQNTLEINND-----NPKIISFNFEFLVYTLRDTIKMNDKLTGIN-NTT 217
  |||
DB 555 TDDQTDSDNDVYLGDMGNLTKEKIMSYNAFFMMVHDLIDSIRKDEHGRICINKDKG 614
  |||
QY 218 HCIDECNRNCLCFDRWVKQKEEWSIKKLFKKKNIOQSYYSNINLEFGEFFVMDKL 277
  |||
DB 615 FTICGCKKRCICFQKWEKKTKEWGIKIDHFRKQKIDPQWTHD-----DFLQTL--L 665
  |||
QY 278 DKDEAKKLENIKRNKNEFSNLE-----NNRDYL-----ENAIELL 315
  |||
DB 666 MKD--LLEETIDTYGDANEIKRIEALLEQAGVGIDFPAALGLTYKGVAKDITITDKL 723
  |||
QY 316 LDHLEFATICKDNTNEAC 335
  |||
DB 724 LQHEQKADKCLKTHTDTC 743
  |||
```

Search completed: April 28, 2003, 10:30:04
Job time : 17.5249 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 6.38122 Seconds
(without alignments)
5197.501 Million cell updates/sec

Title: US-10-087-013-2-COPY_403_747

Perfect score: 1918
Sequence: 1 YEKIQSYLSNDNKFVNIN.....CKDNTNEACETSHNATNP 345

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Minimum number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	771	40.2	3006	2 T28625	variant-specific s
2	501.5	26.1	2647	2 T28161	hypothetical prote
3	499.5	26.0	2212	2 T28157	erythrocyte membra
4	457	23.8	2664	2 T28626	variant-specific s
5	446	23.3	2228	2 T14029	variant-specific s
6	437	22.8	2026	2 T28431	variant surface pr
7	432.5	22.5	2197	2 B71600	variant-specific s
8	403.5	21.0	2706	2 T28155	variant-specific s
9	394.5	20.6	3078	2 T28432	variant-specific s
10	380	19.8	2182	2 T28634	variant-specific s
11	363	18.9	1711	2 C71625	variant-specific s
12	361.5	18.8	2135	2 T14602	variant-specific s
13	357.5	18.6	1729	2 T18396	erythrocyte membra
14	357.5	18.6	2924	2 T18378	variant-specific s
15	331	17.3	2042	2 T18399	variant-specific s
16	162	8.4	807	2 B71605	hypothetical prote
17	155.5	8.1	2265	2 T28677	thoptry protein-
18	153.5	8.0	686	2 G71620	hypothetical prote
19	150.5	7.8	2380	2 E71604	hypothetical prote
20	150	7.8	4550	2 T18440	hypothetical prote
21	149	7.8	1005	2 A64465	hypothetical prote
22	148.5	7.7	3394	2 T18501	hypothetical prote
23	148.5	7.7	3724	2 T18427	hypothetical prote
24	147.5	7.7	1516	2 E71619	RAD2 endonuclease
25	146	7.6	1939	2 T18372	repeat organellar
26	146	7.6	2339	2 A45597	DNA-directed RNA p
27	145.5	7.5	407	2 C45600	asparagine-rich bl
28	144.5	7.5	1979	2 C71622	hypothetical prote
29	143.5	7.5	1166	2 H71609	hypothetical prote

30	142.5	7.4	1247	2 E71616	hypothetical prote
31	142	7.4	1711	2 T18429	hypothetical prote
32	140.5	7.3	1817	2 H71611	probable secreted
33	140	7.3	2136	2 A05037	hypothetical prote
34	139	7.2	2401	2 T28676	hypothetical prote
35	138	7.2	1127	2 T28317	thoptry protein-
36	137	7.1	1435	2 A37793	ORF MSV156 hypote
37	137	7.1	2010	2 B71616	erythrocyte-bindin
38	137	7.1	3973	2 B71612	phosphatase (acid
39	136	7.1	508	2 E71620	hypothetical prote
40	136	7.1	911	2 T18451	hypothetical prote
41	136	7.1	1215	2 B48281	cytochrome associat
42	136	7.1	4961	2 T18489	hypothetical prote
43	135.5	7.1	1817	2 D71606	hypothetical prote
44	135.5	7.1	2657	2 T18497	hypothetical prote
45	135	7.0	2510	2 T28160	hypothetical prote

ALIGNMENTS

RESULT 1
T28625
variant-specific surface protein 3 - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C/Accession: T28625
R/Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson, Cell 82, 89-100, 1995
A>Title: The large diverse gene family var encodes proteins involved in cytoadherence
A:Reference number: Z20487; MIMD:95350813; PMID:7606788
A:Accession: T28625
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3006 <SU>
A:Cross-references: EMBL:L40609; NID:9886376; PID:9886377; PIDN:AAA75397.1
C:Gene: var-3
A:Gene: var-3
A:Introns: 2597/3

Query Match	Score	771	DB 2:	Length	3006:
Best Local Similarity	41.5%				
Matches 147; Conservative 72; Mismatches 115; Indels 20; Gaps 6;					
QY 1 YEKIQSYLSNDNKFVNINSEYKQFYETKQYATNDPFLNLNEGKYCKGLPGEK 60					
DB 392 YTKIEIKKSSSDKSNISNKKYNEFNPKKEVETLQFLKLKKKATCCQKIEEE 451					
QY 61 DITFTNSADKGIIFYRSEYQVCPDCGVKCDGKIKYTHKSDRER-----VNEDYKP 113					
DB 452 VIDE--NKDEDMVFHRSEYCPQPCDVVQCKGKCTEDKKRNDKRSKIRKIIQSEE-- 506					
QY 114 PMGKPTNITVLVSGNKGITOKLEFNCSSTYKKNOKMKCYKDKENINCKLEON 173					
DB 507 -----PTEIHVLSDDKQGITTKLEVFCSSTYBERNVQMKYKNSDYNCENIS 561					
QY 174 TEINNDPKIISFANFELWVYLLRPDIKWNKDKLKTICINN--TTICIDECNRNCLCFDR 232					
DB 562 SYKSDTANVWLSEVCHSAKNLLIDITIKWEHQLKNCINNNTVTCESCKIKCEYER 621					
QY 233 WKQKEEWSIKKLFKKKNIOQSYISNINLFEQFFKVMKLDKDE-AKKKELMENT 291					
DB 622 WIKRKEHEWKVKVNFNNRMSYIYNNISRYVDSFLFQVMEFLDDDEKGMQGFEDL 681					
QY 292 KKKKESNLENNNDYLENAIELLDLKLKATICKDNNTNEACETSHNATNP 345					
DB 682 K-KKEFSKTNTPYKGSQDAIEFLDLKDALTCRDNNSNESCVSKKYKTNP 734					
RESULT 2					
T28161					
hypothetical protein FC93-var11-1 - malaria parasite (Plasmodium falciparum) (fragment)					
C/Species: Plasmodium falciparum					

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28161
R: Hernandez-Rivas, R.; Maltel, D.; Sterkers, Y.; Peterson, D.S.; Wellens, T.E.; Scherf, M.O. Cell. Biol. 17, 604-611, 1997
A>Title: Expressed var genes are found in Plasmodium falciparum subtelomeric regions.
A:Reference number: Z20483; MUID:97154495; PMID:9001213
A:Accession: T28161
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2647 <HER>
A:Cross-references: EMBL:U67959; NID:91794255; PID:91809295; PIDN:AA047438.1
A:Experimental source: strain FCQ27/PNG
C:Genetics:
A:introns: 2158/3
A>Note: FCR3-varT11-1

Query Match 26.1%; Score 501.5; DB 2; Length 2647;
Best Local Similarity 34.4%; Pred. No. 8e-23;
Matches 116; Conservative 54; Mismatches 150; Indels 17; Gaps 11;

2 EKEIOSYLSNDKFNVNINSEYKQFEKLEKTOYATNDPFLNLNKGKCGGLPGEKD 61
Db 395 ENELIRNNESSQSPKNNNNYETDFGNNKKDYQSMND-FLKLNSETPCTNIIDAKSK 453
QY 62 ITFTNSADDKGIFRSEYQVCPDGGYK--CDGKTYTHKSDNRERNNNDYKPPMGVPR 119
Db 454 IDFKR--DPEFTSHTEYCDPCFWCGKLTQADG-TWKRLYENDPQCPKFKYEPKGVPR 510
QY 120 TNLTVLSNGEODITQKLEFNCSSTNYKDKNNQKWEYKDEINRCKLEQNTLEIND 179
Db 511 TEIDVLTGKENDDIYKLEFECKTDGNTGPK-NEEWNCYQVGN-DKCVLENGEELGSE 568
QY 180 NPKIISHNFELWYVLYLRDTIKNDKLTCTCINNTTTHIDECNRCCLCFDWWYQKEE 239
Db 569 K-KVKDYNFLMFWVAHMLDSEMSKSLNCLSKDCTTCNCNCCQCYDKWIKKKYV 627
QY 240 EWSIKKLFYKKNKNIQ-QSYYSNINLFEQ-YFEKVMKLDKDEAKWELMENIKRKNNE 297
Db 628 HWTQIKKHFDKQDFQGWGIFYLETVLEGGDFFTDITAYGDAREIVHIOELQKKKQ 687
QY 298 FSNLENNRDYLENAIELLDHLEKETATTC---KDN 330
Db 688 V-LHEDASNMKTIIDELDLHELKAKOCIVNKHNN 722

RESULT 3
T28157
Erythrocyte membrane protein 1 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
A:Accession: T28157
R: Yang, S.Q.; Wollish, W.S.; Gut, J.; Wu, J.; Ahn, J.; Petersen, C.; Fujio, H.; Akawa
A:Description: The molecular cloning and DNA sequence analysis of Plasmodium falciparum
A:Reference number: Z20479
A:Accession: T28157
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2212 <YAN>
A:Cross-references: EMBL:U31083; NID:91517813; PID:91517814; PIDN:AA06961.1
C:Genetics:
A:Gene: EMP1

Query Match 26.0%; Score 499.5; DB 2; Length 2212;
Best Local Similarity 34.4%; Pred. No. 8.8e-23;
Matches 116; Conservative 56; Mismatches 148; Indels 17; Gaps 11;

QY 2 EKEIOSYLSNDKFNVNINSEYKQFEKLEKTOYATNDPFLNLNKGKCGGLPGEKD 61
Db 395 ENELIRNNESSQSPKNNNNYETDFGNNKKDYQSMND-FLKLNSETPCTNIIDAKSK 453
QY 62 ITFTNSADDKGIFRSEYQVCPDGGYK--CDGKTYTHKSDNRERNNNDYKPPMGVPR 119
Db 395 ENELIRNNESSQSPKNNNNYETDFGNNKKDYQSMND-FLKLNSETPCTNIIDAKSK 453

Db 454 IDFKR--DPEFTSHTEYCDPCFWCGKLTQADG-TWKRLYENDPQCPKFKYEPKGVPR 510
QY 120 TNLTVLSNGEODITQKLEFNCSSTNYKDKNNQKWEYKDEINRCKLEQNTLEIND 179
Db 511 TEIDVLTGKENDDIYKLEFECKTDGNTGPK-NEEWNCYQVGN-DKCVLENGEELGSE 568
QY 180 NPKIISHNFELWYVLYLRDTIKNDKLTCTCINNTTTHIDECNRCCLCFDWWYQKEE 239
Db 569 K-KVKDYNFLMFWVAHMLDSEMSKSLNCLSKDCTTCNCNCCQCYDKWIKKKYV 627
QY 240 EWSIKKLFYKKNKNIQ-QSYYSNINLFEQ-YFEKVMKLDKDEAKWELMENIKRKNNE 297
Db 628 HWTQIKKHFDKQDFQGWGIFYLETVLEGGDFFTDITAYGDAREIVHIOELQKKKQ 687
QY 298 FSNLENNRDYLENAIELLDHLEKETATTC---KDN 330
Db 688 V-LHEDASNMKTIIDELDLHELKAKOCIVNKHNN 722

RESULT 4
T28626
variant-specific surface protein 2 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
A:Accession: T28626
R: Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson,
A>Title: The large diverse gene family var encodes proteins involved in cytoadherence
A:Reference number: Z20487; MUID:95330813; PMID:7606788
A:Accession: T28626
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2664 <SU>
A:Cross-references: EMBL:L40609; NID:9886376; PID:9886378; PIDN:AA75398.1
C:Genetics:
A:introns: 2197/3
A>Note: var-2

Query Match 23.8%; Score 457; DB 2; Length 2664;
Best Local Similarity 31.1%; Pred. No. 4.3e-20;
Matches 114; Conservative 51; Mismatches 153; Indels 48; Gaps 9;

QY 2 EKEIOSYLSNDKFNVNINSEYKQFEKLEKTOYATNDPFLNLNKGKCGGLPGEKD 53
Db 392 ENELIRNNESSQSPKNNNNYETDFGNNKKDYQSMND-FLKLNSETPCTNIIDAKSK 451
QY 54 -----GLPGEKDTFTNSADDKGIFRSEYQVCPDGGYKCDGKTYTHKSDNDRERY 106
Db 452 IDFKTVQGGGAGGAGSADSNNSNKTFSHQYCEPCGGGVELIGNEW--KKNKGECK 509
QY 107 NNEIYKPPMGVPRNTITVLSNGEODITQKLEFNCSSTNYKDKNNQKWEYKDEINRCKLEQNTLEIND 166
Db 510 GKKRYNIPKGNHVIPLSFGDEHKEIKIDFQCAESMSDSKLTQKCYGKEYE 569
QY 167 RCKLEQNTLEINDNPKT-ISPHEFELWYVLYLRDTIKNDKLTCTCINNTT-HCIDECN 224
Db 570 VCTLENNKSEEDPEEIOKTFHNFYFWIRHLNDSIEMWDMKINCIERAKKCKNECK 629
QY 225 RNCLEDRWYKQKEEEMNSIKKLEFKKKNIOQSYYSNINLFEYGEFYKMDKIDQDEAKW 284
Db 630 TDCGCFQRIWIKKKEWGEIKKFK-----TQGFSEIFGN--NYDFYLENVLND---- 677
QY 285 KEMLENNKKNFSLNENRDYL-----ENATLELDHLEKETATTCQDN 330
Db 678 -ELFQDITAYGNSQRIQIKDTLAKKTQADDATEQKNTIDLLFEYSEAEKCK-K 734
QY 331 TNEACE 336
Db 735 IOBEQ 740

RESULT 5
T14029

variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
 N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)
 C:Species: Plasmodium falciparum
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
 C:Accession: T14029
 R:Chen, Q.; Barragan, A.; Fernandez, V.; Sundstrom, A.; Schlichtherle, M.; Sahlen, A.; C
 J. Exp. Med. 187, 15-23, 1998
 A:Title: Identification of Plasmodium falciparum erythrocyte membrane protein 1 (PEMP1)
 A:Reference number: 217860; MID:98080592; PMID:9419207
 A:Accession: T14029
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2228 <CHE>
 A:Cross-references: EMBL:AF003473; NID:g3961467; PID:g3961468; PIDN:AAC05730.1
 C:Genetics:
 A:Note: FCR351.2-vari

Query Match 23.3%; Score 446; DB 2; Length 2228;
 Best Local Similarity 31.4%; Pred. No. 1.7e-19;
 Matches 131; Conservative 53; Mismatches 133; Indels 100; Gaps 20;
 1 YEKIQLSYLSNDKRFVNNIN--SEYKQFYERKLETOYATNDFLNLNKGKCK-----53
 DB 415 YTEKIDSGGGRKRAVGGITTKYEGYERKLEKNDGYDAFLGLLNKCKCKITDGG 474
 QY 54 -----GGLPEKIDITFTNSADDKGIFRSEYCOVPCDGVCKDGKIKYTH 97
 DB 475 GKINFRVNSGGVGGSGCTSGASGTND-EKKGTFFYRSEYCOVPCDGVCK-----H 526
 QY 98 KSNDRERANNEDYKRPWG--VAPTN---ITVLYSGNEGD---ITQKLENEC---NSS- 145
 DB 527 KGNQWER-KTKYKRMKSKLYRPNKMWLLKSLKRVKDMILKMKWEKFCITLONSSD 585
 QY 146 -----FMYKRNKNOK-----WECYQKDEMINCKLEPONTINND-----179
 DB 586 GSVGVYTTTGAAGSGNSEKKELYDEMKC-YKHNEVQKVVNNGVEYEDDEKAGAGCITLP 644
 QY 180 NPR-----IISFHNFEFLWVYLLRDTIKMNDK-LKTCINN-TTTHCI 220
 DB 645 NPKKNEVSEAKSQNNHADIQKTFHDFEYVVAHMLKDSIHMTRKRLKSCISDGKTKCR 704
 QY 221 DEGNRNCICFDRVVKQKEEEMNSIKILFTKKKNIQOSY-SNINNLEGEFFR-----V 273
 DB 705 NGSKKCKDCDEKRVKQKEEMKPKIKHFKTQESIPGTYFTTELILKLOFLEDEEENT 764
 QY 274 MDKIDKDEAKWELMENIKRKNFESNLE--NNRDYLENAIELLDHLKFTATCKD 328
 DB 765 ENSLDAEAEELKHLKIKLENE-NMLAVVNVNGTEKTLMDKLNHLEMDATKCKD 820

RESULT 6
 T28431
 Variant surface protein 1 homolog AA4VAR - malaria parasite (Plasmodium falciparum) (frag
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T28431
 R:Smith, J.D.; Kyes, S.; Craigh, A.G.; Fagan, T.; Hudson-Taylor, D.; Miller, L.H.; Baruch
 Mol. Biochem. Parasitol. 97, 133-148, 1998
 A:Title: Analysis of adhesive domains from the AA4VAR Plasmodium falciparum erythrocyte m
 A:Reference number: Z20486; MID:99094502; PMID:9879893
 A:Accession: T28431
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3026 <SMI>
 A:Cross-references: EMBL:L42244; NID:g3540144; PID:g3540145; PIDN:AAD03351.1
 C:Genetics:
 A:Gene: var
 A:Introns: 2906/3

Query Match 22.8%; Score 437; DB 2; Length 3026;
 Best Local Similarity 27.9%; Pred. No. 8.2e-19;
 Matches 112; Conservative 73; Mismatches 139; Indels 78; Gaps 14;

QY 1 YEKIQLSYL----SNDKRV-----NNINSEYKQFYERKLETOYATNDFLNLNKGKCK 52
 DB 394 FEKOKKKYTFDEIKKNDTITTEYGTINNNYRRDFQKYLEE-KYKTVDAFLNLNKEK 452
 QY 53 KG-----GLPEKIDITFTNSADDKGIFRSEYCOVPCDGVCKDGKIKYTHKSDNDREVN 108
 DB 453 KNPVEGGRKKYIDNDNE--TFSHTYRRCPCMGIE-----EOKDKWRKIND 502
 QY 109 -----EDKPPMGVFPYITVLYSGNEGDITQKLENSSTYVKKDN-----153
 DB 503 HSAKKEELATPKENKKTITNVLTSGEHEDLAKRLKECTCTQONCGSDGCGSDS 562
 QY 154 ---OKWECYKDE-----NINCKLEQNTLEINNDPKIISPHNF 190
 DB 563 SLCEPWQCYQPDQLEKVGGEVDKLLKAGAGLCIFEMKKEKRYK-----KQTFNNF 616
 QY 191 ELWVYLLRDTIKNDKTKCIN-NTTTHCIDCNRCICFDRVVKQKEEEMNSIKILFT 249
 DB 617 NFWVAHLKSDIKWKTOLKLSLSEDKLKCEKCKSCSECFKWKIEKKEWIKRVADQFN 676
 QY 250 KKKK-IOOSYYSNINLFESEYFPKVMKLDKDEAKWELMENIKRKNFESNLENNRDYL 308
 DB 677 KQNDPFLMKRYLVLETLNLYFENTOKAYGDLKSIQEMKMKIKENKQNNRTKDD-----732
 QY 309 ENAIELLDHLKETATC-----KDNNTNEACETSHNATNP 345
 DB 733 EDALDVLFDHEKEAECDLIDHEDDDDDCEVEIEKIPNPP 774

RESULT 7
 B71600
 variant-specific surface protein 1 homolog PFBI055c - malaria parasite (Plasmodium fa
 N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)
 C:Species: Plasmodium falciparum
 C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C:Accession: B71600
 R:Carder, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
 ; Perle, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; MID:99021743; PMID:9804551
 A:Accession: B71600
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2197 <GAR>
 A:Cross-references: GB:AE001434; GB:AE001362; NID:g3845341; PIDN:AAC71996.1; PID:g384
 C:Experimental source: clone 3D7
 C:Genetics: PFBI055c

Query Match 22.5%; Score 432.5; DB 2; Length 2197;
 Best Local Similarity 29.4%; Pred. No. 1.1e-18;
 Matches 116; Conservative 62; Mismatches 126; Indels 91; Gaps 14;
 23 YKQFYERKLETOYATNDFLNLNKGKCKGGLPEKIDITFTNSAD---DKGIFRSE 78
 DB 443 YEKKFEKLESESEGVDDFLKLNEDVCKIKIDEKEKIDFTPADKSNNEGTFFHSE 502
 QY 79 YCQVCPGCGKCGKICIKTHSDNDRVNNEDYRPPGVPTNTVLYSGNEGDITQKL 138
 DB 503 YKPCPCGKVRKRDNDKDYDKCTR--GLVEPASGAGTPIKILSGKQKIEITKL 560
 QY 139 ENFCN-----SSTN-----YKDKNNOK-----WECYK-----DE 163
 DB 561 KAFCDQNGDTNSVANGAGDGSNSKSELYEEMKCKNEVQKVDKNGEEDDEE 620
 QY 164 NIN-----CKLE-----QNTLEINNDPKIISFHNFEFLWVYLLRDTIKMNDKLT 210
 DB 621 DVDKVRKAGGICILENKKHSRNNSNEPEQFTFDFEFYFWIGRLNDSMYRGLVNS 680
 QY 211 CINN-TTTHCIDCNRCICFDRVVKQKEEEMNSIKILFTKKKNIQOSYYSNINLPEGY 269
 DB 681 CINNPKRKRCKCKDCGCFKEMIGKKEEMNIKHFITQEAFFKKNRNSGIDMFSG- 739

270 FEKVADKID-----KDEAKKELMENIRKKNEFSNL----- 301
 Db 740 ---LMDSDVVELLELEQLQFDIKDYGDKELKIGIKELLEDKRRKKQAEAVVYVA 796
 QY 302 ENNRDYLENALIELLDHKETATICKDNNTNEACE 336
 Db 797 DNQK---KTTIDKLQHEGDDANCLKTH-KEKCE 827
 A:Gene: var-1
 A:Introns: 2611/3

RESULT 8

T2815
 variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum) (fragments)
 N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T2815
 R:Rove, J.A.; Moulds, J.M.; Newbold, C.I.; Miller, L.H.
 Nature 388, 292-295, 1997
 Title: Plasmodium falciparum rosetting is mediated by PfEMP1 and requires complement
 Reference number: 220477; MUID:97373957; PMID:9230440
 C:Accession: T2815
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2706 <ROW>
 A:Cross-references: EMBL:Y13402; PIDN:CAA73831.1
 A:Experimental source: strain IT 4/25/5
 C:Genetics:
 A:Introns: 2493/3
 A:Note: R29R+var1

Query Match 21.0%; Score 403.5; DB 2; Length 2706;
 Best Local Similarity 29.1%; Pred. No. 8.2e-17;
 Matches 104; Conservative 69; Mismatches 142; Indels 43; Gaps 14;

QY 1 YKEEIOSYLSNNKFNNT--NSEYYKOFYEKLEKQYATNDPFLNLNGYCKGGSLPGE 59
 Db 384 FPKOKRKTH-NEIKIKITNISNNENDEKFEYENDDKGYSTINTFLESLNGKOCODNIDKK 442
 QY 60 KDITFTNSADDKGIFYSRSEYQVPCDQGVKDGKIKTHKSDNDRERNEDYKPPW---- 115
 Db 443 NKTNFKNNLE---TREGPSYCEPCPIYGVKCSNEKCTPTENEMW---NSNRLPTDST 495
 QY 116 -GVKPTNITVYSGNEQGITOKLENFCSSTNYKDNKNQKNECYKDEINRCKLEQNT 174
 Db 496 KMLNMTNIDMLVNDGIGNAIDNELEKNCYKILGKIKQKQCOYL-NIDQCKI--NN 552
 QY 175 EINN--DNPKIISFNPELWYTLRDTIKMNDKLTCTI--NNTTTHCIDEENRCL 228
 Db 553 VMSGTFDNK--IARVLFQRMIRYVRDHNRLKEKIDVICIKENINENICIKCTNCE 610
 QY 229 CEDRWYKQKEEWSIKKLTFTKKNIQ-----QSYSNINLPEGYFPVMDKLDK 280
 Db 611 CVGKMLEKKEAWDKINQYKQKNIHMFILPIWITGEFEKIT--FPNDFKALDEDVT- 667
 QY 281 EAKKMEIMNIRKKRKEFSLENRDYLENAIEFLDHLKERTATICK---DNNTNAC 335
 Db 668 ---INVLDLTKECODTHCKIKIRSIDVLDLKEIISWLNKNIKLEVCSSHDEKHEYC 721
 A:Status: preliminary; translated from GB/EMBL/DBJ

RESULT 9

T28432
 variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
 N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T28432
 R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herfeltdt, J.A.; Peterson, D.S.
 Cell 82, 89-100, 1995
 Title: The large diverse gene family var encodes proteins involved in cytoadherence an
 A:Reference number: 220487; MUID:95330813; PMID:7606788
 C:Accession: T28432
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-3078 <SUX>
 A:Cross-references: EMBL:L40608; NID:9886374; PID:9886375; PIDN:AAA75396.1
 C:Genetics:
 A:Gene: var-1
 A:Introns: 2611/3

Query Match 20.6%; Score 394.5; DB 2; Length 3078;
 Best Local Similarity 26.2%; Pred. No. 3.3e-16;
 Matches 122; Conservative 76; Mismatches 135; Indels 133; Gaps 20;

QY 1 YKEIQSY-----LSNDKFNVNINSE-YIQOFYEKLEKQYATNDPFLNLNGK 50
 Db 394 YDEIKKYENGASGSGSRKRAAGTITTYDYEKFFDELNKSERYVDKFLKLSNEE 453
 QY 51 VC-----KGLPGKIDITFNSA-----DDKGFYSRSEYQVPCDQGVKDGKIKTH 97
 Db 454 ITTKYKDEGGITDRKNVNSDTSAGASGVNVSQGTFRSKTCQCPYCGVK---KVN 509
 QY 98 -KSDNDRERNV---EDYKPPGVKPTNITVYSGNEQGITOKLENFCSN----- 143
 Db 510 GSSSNWEKKNKCKCKSGKLYEPKPDKEGTITITLISGKHDDIREKLNKFCDEKNPDTI 569
 QY 144 ---SSTNKKDKNN-----QKMECYKDEINRCKLEQNTENNDPK-----IT-- 184
 Db 570 NSGSGTGGSGGSGSGROELYEEMKC-YKGEVYVVGHEDEDEYENYKNAAGLICITKN 628
 QY 185 -----SFHNFPELWYTLRDTIKMNDKLTCTI--NNTTTHC-ID 221
 Db 629 QKKNEEGNTSEKPEDEIQTFNPFYVVAHMLKDSITHMKKTLQRLQONRIRKCGNN 688
 QY 222 EENRCLCPDRWYKQKEEWSIKKLTFTKKNIQ-----QSYSNINLPE 267
 Db 689 KKNNDCECFKRWITOKKDEMGKIVOHF-KTONIKRGSGSDNTALIPDHDVLOYNLOE 747
 QY 268 GVF-----FKWMDKDKDEAKKMEINIRKKNEFSNLENNRD-----YLEN 310
 Db 748 EFLKQDSEASEKSENSLDAEA-----ELKHLREISEDNQKESVGGVTEQKN 801
 QY 311 AIELLDHLKETATIC-----KDNNTNEACETSHNATNP 345
 Db 802 IMDKILNTEKDEADICLETIHEDEEKEKGDGNECTIEGERNRNP 847
 A:Status: preliminary; translated from GB/EMBL/DBJ

RESULT 10

T28634
 variant-specific surface protein 7 - malaria parasite (Plasmodium falciparum)

N:Alternate names: erythrocyte membrane binding protein 7 (EMP7)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T28634
 R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herfeltdt, J.A.; Peterson,
 Cell 82, 89-100, 1995
 Title: The large diverse gene family var encodes proteins involved in cytoadherence
 A:Reference number: 220487; MUID:95330813; PMID:7606788
 C:Accession: T28634
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-2182 <SOX>
 A:Cross-references: EMBL:L42636; NID:9886379; PID:9886380; PIDN:AAA75399.1
 C:Genetics:
 A:Note: var-7

Query Match 19.8%; Score 380; DB 2; Length 2182;
 Best Local Similarity 28.7%; Pred. No. 1.8e-15;
 Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

QY 1 YKEI-----QSYLSNDKFNVNINSEYQFYEKLEKQYATNDPFL 43
 Db 413 YETELISGSGSGSPKRTKRAASSSSSDN-----GYEKFRTKLEKGYQVDKFL 464
 QY 44 NLNKGKCKG---GLPGEKIDITFNSADDKGIFYSRSEYQVPCDQGVKDGKIKTH 99
 Db 465 KILNKRGICQKQPOYGNENKADNVDFNKKYK-TSRTETICPCPCWGLKKGKGGPPMKVKG 523

QY 100 DNDREKVNEDYPPWGVKPTNTVLYSGNEGDDITOKLENFCSSTNYKDNKNNQKWEY 159
 DB 524 DKTGSAKRTYDPK---NTTDPVLYPDKSOONILTKKKNFCEKGNP--GGGQIKKQCY 579
 QY 160 YKD-----ENINCKLEQNTNNDNPKTISFNFELWVLYLTPDKMDLKTICNN 214
 DB 580 YDEHRSSKNNNNCEGTWDKFTQGGKQYKASVYFWDVMDVMDVSEKTELSTICINN 639
 QY 215 TTT---HCIDECNRKCLCFDRVWKQKEEEMNSIKLFTKKNKIOOSYYSINNLEGEYF 270
 DB 640 NTNGNCRNNKCKTDCGGCQKQVKEKKQOEWMALIKDFGQGTIYQO-----KGLYFSP 694
 QY 271 EKVMKLDKDEAKWELKEMIRKKNKESLNNRDLF-----NALELL 316
 DB 695 YGVLDLY---LKGMLQWIKQVHSDTIDIKIKILLDEDAVAAYLGKNDWTIDKIL 750
 QY 317 DHKETATICKONNTNEACE 336
 DB 751 QHEKQAEQCKK--QEECE 768

RESULT 11

variant-specific surface protein 1 homolog PFB0010w - malaria parasite (Plasmodium falci-
 C:Species: Plasmodium falciparum
 C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C:Accession: C71625
 R:Gardner, M.J.; Tetteh, H.; Garnuci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; M01D:99021743; PMID:9804551
 A:Accession: C71625
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1711 <GAP>
 A:Cross-references: GB:AE001366; GB:AE001362; NID:93845070; PIDN:AACT1792.1; PID:9384507
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PFB0010w

Query Match 18.9%; Score 363; DB 2; Length 1711;
 Best Local Similarity 27.5%; Pred. No. 1.5e-14;
 Matches 100; Conservative 60; Mismatches 130; Indels 74; Gaps 14;

QY 1 YEKELSYLSNDKFNVNINSEYKQYERKLEQVATNDTFLNLNEGYCKGLP-GE 59
 DB 424 YEHFNEELNEKQKDR-----SFLQLSKERICKERIOVGE 460
 QY 60 KDITFTNSADKGIYRSEYQVCPDGVKCDGIKYTHKSDND--REKVNEDYKPPWGVK 118
 DB 461 ETANYGNFENESMTFSHTEKCDRCPLCGVDCSSDNCKRDKSGDEITIKYRPE---N 517
 QY 119 PTNITVLYSGNEGDDITOKLENFCSSTNYKKNKNN---QKMECY---KDE---NINR 167
 DB 518 TTKIPKLTAKKRTGILKTKKFKCKNS---DGNNGQIKKWKCHYKKNKDKDNGDINN 573
 QY 168 C-KLEQNTNNDNPKTISFNFELWVLYLTPDKMDLKTICINNNTT-HCIDECNR 225
 DB 574 CIGDQMTSKNYYP--ISYSPFYGTIDMLNLSIEKRLSCINDAKLGCKRCKGN 631
 QY 226 NCICFDNRVQKEEEMNSIKLFTKKNKIOOSYYS--NINLFEYFPPVMDKLDKDEAKW 284
 DB 632 PCCEYKAWVEKKDEMDKIEFFRKQKDLKAGMDAGLEYLENI-----680
 QY 285 KEMLEKTKKKNKESLNNRDLLENAIELLDHLKETAII-----CKDNTT 331
 DB 681 --FLEDDKRNAGDPKYLEKEKILGKENEVODPLKTKTIDTLEKELNEAKKVCENP 738
 QY 332 NEAC 335
 DB 739 DNEC 742

RESULT 12

variant-specific surface protein - malaria parasite (Plasmodium falciparum) (fragment
 C:Species: Plasmodium falciparum
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
 C:Accession: T14602
 R:Voss, T.S.; Felger, I.; Welas, N.; Beck, H.P.
 submitted to the EMBL Data Library, February 1998
 A:Description: Identification of a conserved 5' flanking region of Plasmodium falcipa-
 A:Reference number: Z18158
 A:Accession: T14602
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2135 <VOS>
 A:Cross-references: EMBL:AF050740; NID:92944094; PID:92944095; PIDN:AAOC5220.1
 C:Genetics:
 A:Gene: varph17

Query Match 18.8%; Score 361.5; DB 2; Length 2135;
 Best Local Similarity 27.5%; Pred. No. 2.4e-14;
 Matches 111; Conservative 57; Mismatches 142; Indels 93; Gaps 16;

QY 1 YEKELSYLSNDKFNVNINSEYKQYERKLEQVATNDTFLNLNEGYCKGLP-GE 53
 DB 401 YGEI-SLKRKRKRDAGISTKYDGEKRYEKLK-SEYRVGEFLGLNNETCKEVK 458
 QY 54 GGLPGEKIDITFTNSADKGIYRSEYQVCPDGVKCDGIKYTHKSDNDREKVN 108
 DB 459 DDEKGIKDEFTVNSGASGDVKNKTFYRTYCECPGCAEORNGVGMKARDKDCSPG 518
 QY 109 EDYKPPWGVKPTNTVLYSGNEGDDITOKLENF-----142
 DB 519 NDYTK---YKKKEIPILTGDKTSEIYERKCKKNGKANGREGVSGSENGAASNS 575
 QY 143 -NSTNAYKKNKNN---QKMECY---KDN-----INCKLEQNTNNDNPKTISFHN 188
 DB 576 DNATGTGCGGNSDSSLCSEKFWTCYKRYKKNENKNGKIDNFC--VQAGMNSKKDKVKSNA 634
 QY 189 FPELWVLYLTPDKMDLKTICINN--NTTHCIDECNRKCLCFDRVWKQKEEEMNSIKK 246
 DB 635 FFDWMDHMDLIDSKMNEKGCINKDNGKT-CIKGCKSKDCGLKAWQKKEWKLILE 693
 QY 247 LFTKKNKIOOSYYSINNLEGEYFVMDKLDKDEAKWELKEMIRKKNKESLNNRDL 306
 DB 694 HFTMGGFGDGEHQRLGFTHDVLYNLTKD-----KELKLI---EETGNTETTKH 742
 QY 307 YLE-----NALELLDHKETATICKD 328
 DB 743 IKELDKREKRDAGTGVAAGTGPRNIMDKLIEHLEQAKKCKD 785

RESULT 13

erythrocyte membrane protein variant 2 - malaria parasite (Plasmodium falciparum) (fr
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T18396
 R:Baruch, D.I.; Pasloske, B.L.; Singh, H.B.; Bi, X.; Ma, X.C.; Feldman, M.; Taraschl,
 Cell 82, 77-87, 1995
 A:Title: Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen a
 A:Reference number: Z18925; M01D:95330812; PMID:7541722
 A:Accession: T18396
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1729 <BAR>
 A:Cross-references: EMBL:U027339; NID:9914920; PID:9914921; PIDN:AAA89134.1
 C:Genetics:
 A:Gene: EMP1
 A:Note: var-2

Query Match 18.6%; Score 357.5; DB 2; Length 1729;

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:46 ; Search time 3.31492 Seconds

(without alignments)
4316.649 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747

Perfect score: 1918
Sequence: 1 YKEIQSYLSNDNKFVNIN.....CKDNTNACETSHNATNP 345

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

al number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	149	7.8	1005	1	RASO_METUA
2	146	7.6	2339	1	RPCI_PLAFA
3	140	7.3	2136	1	YCF2_MARPO
4	137	7.1	1435	1	EBAL_PLAFC
5	136	7.1	1147	1	CGAL_HELPI
6	130	6.8	1956	1	ATX1_PLAFA
7	127.5	6.6	1790	1	USO1_YEAST
8	122	6.4	1805	1	HMW2_MYCGE
9	121	6.3	1153	1	PVDB_PLAKN
10	120.5	6.3	537	1	ARP_PLAFA
11	119.5	6.2	782	1	Y091_CAEBL
12	119	6.2	971	1	Y228_BORBU
13	119	6.2	1073	1	PVDA_PLAKN
14	118.5	6.2	1085	1	IFH1_YEAST
15	118.5	6.2	1220	1	IF2E_HUMAN
16	118	6.2	1070	1	PVDR_PLAFA
17	117	6.1	500	1	RT03_PROWI
18	116	6.0	1130	1	YL17_CAEBL
19	115.5	6.0	864	1	RASO_SULSO
20	115	6.0	494	1	PRE_STRAG
21	114.5	6.0	1202	1	RPW2_YEAST
22	114.5	6.0	1280	1	KCPG_XENLA
23	114	5.9	1701	1	MSPI_PLAFA
24	113.5	5.9	537	1	Y029_XENLA
25	113.5	5.9	816	1	YG3A_YEAST
26	113	5.9	357	1	TRMA_CAMJE
27	113	5.9	766	1	STRB_YEAST
28	113	5.9	1186	1	CAGA_HELPI
29	112	5.8	709	1	CDAT_PLAFA
30	112	5.8	1036	1	Y414_MYCGE
31	112	5.8	1726	1	MSPI_PLAFA
32	111.5	5.8	964	1	DPOL_CBEPI
33	111.5	5.8	1082	1	SP23_YEAST

34	111.5	5.8	1251	1	RB2_PLAFA
35	111	5.8	1038	1	CIN8_YEAST
36	111	5.8	1726	1	MSPI_PLAFA
37	110.5	5.8	11070	1	PVDC_PLAKN
38	110.5	5.8	1164	1	BAG_STRAG
39	110	5.7	1169	1	EXB_BORBU
40	109.5	5.7	609	1	YSW1_YEAST
41	109.5	5.7	630	1	Y242_MYCGE
42	109	5.7	886	1	RASO_SULSO
43	108.5	5.7	731	1	BAF1_YEAST
44	108	5.6	451	1	ARP2_PLAFA
45	108	5.6	1076	1	RPOB_ASTLO

ALIGNMENTS

RESULT 1
RASO_METUA STANDARD; PRT; 1005 AA.
AC 058718;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE DNA double-strand break repair rad50 AtPase.
GN RAD50 OR M1322.
OS Methanococcus jannaschli.
OC Archaea: Euryarchaeota: Methanococci: Methanococcales;
OC Methanocaldococcaceae: Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.L.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Georgiades N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Uiterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Bordovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT Jannaschli.
RL Science 273:1058-1073(1996).
CC -1- SUBUNIT: Forms a complex with mreII (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -1- FUNCTION: INVOLVED IN DNA double-strand break repair (DSBR). The
CC rad50/mreII complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mreII by unwinding
CC and/or repositioning DNA ends into the mreII active site (By
CC similarity).
CC -1- SUBUNIT: Forms a complex with mreII (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL: U67572; AAB99331.1; .
CC TRIGR; M13322; .
DR InterPro: IPR003439; ABC_transport.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02463; SMC_N; 1.
DR Pfam: PF02483; SMC_C; 1.
DR ProDom: PD000006; ABC_transport; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NF_BIND 32 39 ATP (By similarity).
FT DOMAIN 158 849 COILED COIL (POTENTIAL).

CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: M73770; AAA29729.1; -
DR	InterPro: IPR000722; RNA_pol_A.
DR	InterPro: IPR002879; RNA_pol_A2.
DR	Pfam: PF00623; RNA_pol_A; 1.
DR	Pfam: PF01854; RNA_pol_A2; 2.
KW	transferase; DNA-directed RNA polymerase; Transcription; zinc;
KW	zinc-finger; Nuclear protein.
FT	ZN_FING 88 101 C3H-TYPE (POTENTIAL).
SO	SEQUENCE 2339 AA; 272829 MW; EDD899363086FD48 CRC64;
Query Match 7.6%; Score 146; DB 1; Length 2339;	
Best Local Similarity 20.2%; Pred. No. 0.079;	
Matches 79; Conservative 63; Mismatches 128; Indels 122; Gaps 20;	
QY	9 LSNDKRVNNINSEYKQ--FYEKL-----KEQYATNDFTFLNLNEGKYGCGLPGE 59
DB	1229 LEHDEFYLNKTIWESTDQCNEDTIRSLDNKNVSYIHN-----QGKHLSTLQCAE 1281
QY	60 KDITFTNSADDKGLFRSEYQCVPCDGVCKDGIKYTHKSDNDRVRNNEDEYKPPWGVKP 119
DB	1282 EHITITNTNNDN-----IYVEDIEKELSKMTTKERKQSFKG 1317
QY	120 TITITLYLGSNEGQD--TQKLENF-----CNSSTNYKDKNNQWE-----C- 158
DB	1318 TTRDHEDESEOMKFTYKAKFLEKKKGGMHECNDIEY---NTQYDNIQYNISCN 1374
QY	159 YUKDNIIRCKLEQNT--EINND-----NPKII--SPHNEFLWVLYLADDTIKM--- 205
DB	1375 YIKSQNL-----ENTHHQVANDLSFIKNVYILPRKEHSLFHVNDY--RNVEIENLM 1426
QY	206 DKLKTCINNTTTHCIDECNRCILCFDRWVQKEEWMSSIKKLFTRKKNIOOYSYSNINL 265
DB	1427 DKKIFLSEKENVV-----GQKYNMSKNLKKTEI-----INNI 1462
QY	266 FEGFEFKV--MDKLQDEAKW-----KELMENIKRKNSESLNNRDYLENAIE 313
DB	1463 YRNEKKILNRKTKTMDNDNTWSSDDSIATKATIKKIKKEKKRYHPRKEKEWEDRRNYK 1522
QY	314 LLDD-HLKETATICKDNNTTEACETSHNATTN 344
DB	1523 MITDNNNNNDNNNNNDNNNNNDNNNNNNNNNN 1554
RESULT 3	
YCF2_MARPO	STANDARD; PRT; 2136 AA.
AC	P09975;
DT	01-MAR-1989 (Rel. 10, created)
DT	01-MAR-1989 (Rel. 10, last sequence update)
DT	16-OCT-2001 (Rel. 40, last annotation update)
DE	Hypothetical 259 kDa protein ycf2 (ORF 2136).
GN	YCF2.
OS	Marchantia polymorpha (Liverwort).
OG	Chloroplast.
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC	Marchantiopsida; Marchantiales; Marchantiaceae; Marchantiales;
OC	Marchantiaceae; Marchantia.
OX	NCBI_TaxID=3197;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA	Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA	Ozeki H.;
RT	"Chloroplast gene organization deduced from complete sequence of
RT	liverwort Marchantia polymorpha chloroplast DNA.";
LT	Nature 322:572-574(1986).

```

CC -1- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X04465; CAA28078.1; -
CC PIR: A05037; A05037.
CC PIR: S01591; S01591.
CC InterPro: IPR003959; AAA_ATPase_cent.
CC Pfam: PF00004; AAA_1.
CC Chloplast; Hypothetical protein.
CC SEQUENCE 2136 AA; 25991 MW; 5BD170C9CCF61197 CRC64;

Query Match
Best Local Similarity 23.4%; Score 140; DB 1; Length 2136;
Matches 92; Conservative 51; Mismatches 91; Indels 160; Gaps 24;

QY 4 EI0SYLNDKRY-----NNISEYKQFY-EKLEK-----OYATN 39
DB 508 EIIISINIDKPYINKISLKNKKOKRYLKNKSSDNFRINLMKINKYSSQOFVSN 567
QY 40 DTEL-----NLNKGKCKGGLPGEKDIPTNSAD--KGIFYRSEYQVCPDGVKCDG 92
DB 568 NSFLNPAFEIQQNYLK-----KNILFEKFLNEVSNFFQYK-----KC-- 611
QY 93 IKTHKSDNDREVNNDIKPVGKPTN-----TVYSGNDOGITQKL-ENFCN 143
DB 612 -----KTLNIFLKFASLEKILKRNKKFTSILKFKF-- 644
QY 144 SSTYKDKNNQKQECYKKN-----INRCKLEONTN-----NDPKITSFNFELWYTL 197
DB 645 -----YKNKLNBNQ--YKISQLQNEKEKLNKKRKNFOFNPIKILSFNNSK--KNYI 696
QY 198 LRDITKMDLKTCTINNTTHGIDECNRNCLCFDRWYKQEE-----ENNSIKLETKR 251
DB 697 LQNKRYFN--KMLINNKIL-----TWKIKSNKLVIENSEYN--KILMNK 736
QY 252 KNIQOQSYVINNLKESYFK-----VMDKLDKDEAKKELM-----EN 290
DB 737 KNNKFESEK-NSVLDTEFFNKKSFNITVIEFKLKIQLNFOELKILNCFSLFNSKN 795
QY 291 IKRRK-----NE-----FSNLENNRDYLE 309
DB 796 IKTKIKFKNSYFINENITTFESFNDRKFNIFLE 829

RESULT 4
EBAL_PLAFC
ID EBAL_PLAFC STANDARD; PRT; 1435 AA.
AC P19214;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Erythrocyte-binding antigen EBA-175.
DS Plasmodium falciparum (isolate Camp / Malaysia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5835;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90377299; Pubmed=2204835;
RA Sim B.K.L.;
RT "Sequence conservation of a functional domain of erythrocyte binding
RT antigen 175 in Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 41:293-296(1990).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X52524; CAA36756.1; -
CC PIR: S11561; S11561.
CC Antigen.
CC DOMAIN 159 1104
CC ESSENTIAL FOR BINDING TO
CC ERYTHROCYTES.
FT VARIANT 1031 1031 E->V (IN STRAINS PCR-3 AND ITG).
FT SEQUENCE 1435 AA; 167389 MW; 32A4309021B1C3D6 CRC64;

Query Match
Best Local Similarity 20.5%; Score 137; DB 1; Length 1435;
Matches 86; Conservative 57; Mismatches 149; Indels 128; Gaps 19;

QY 1 YEKEIOSYLSNDN-----KRVN-NINSEYKQFY-EKLEKQYATNDFMLNLNKGKCYK- 53
DB 75 FNKRWKSYGPPNDIKNMSLKNHNNEMFNNTQSEFSTS-----SLIKONKKYVPI 126
QY 54 GGLPGEKDIPTNSADDKGIFYRSEYQVCPDGVKCDGIKYTHKSDNDRR-VNNDYK 112
DB 127 NAYAVSILSFLDSRING--RNTSSNNEVLSNCKEKKRGMKMDCKKKNDSNIVCIPDRR 185
QY 113 PPMGVKPTNITVLYSGNEQD-----ITQKLEN-----FQNSSTN----- 147
DB 186 IQCLIVMLSIKTYTKYTKMDHFIEASKESQLLKKNNDKNYNSKFCNDLKNSLDGYHL 245
QY 148 -----YDKNNOK-----WECYKQD--E 163
DB 246 AMGNMDQEGYSTKAENKIOEVPFAGAGEISEHKIKNFRKRWNEFPEKLEAMLSHKN 305
QY 164 NINRCKLEONTNNDPKIISFNFPFLWYTLRDITKMDLKTCTINNTTH--GIDE 222
DB 306 NINRCKNIPOEL-QITQWIKEMHGER-----LIERNRSKLPSK--CKNNTLYACEKE 358
QY 223 CNRNCLEFDRWYKQEEENWSIKLFTKKNKNIQOQSYSNINNLPEGYEFYVMDILRDEA 282
DB 359 CIDCMKRYWIIISKFEMHTLSKEYETQKVPKEN-----AENYLIKISE--NRNDA 408
QY 283 KWKELMNIRKKNKEFNLENRRDYLENAIELLDLHLEKATICK-----DNTNFACE 336
DB 409 KVSILLNNQAEYSKYD-----CKHTTTLKSVYLVNGNDNTIKERE 450

RESULT 5
CGAL_HELPY
ID CGAL_HELPY STANDARD; PRT; 1147 AA.
AC P80200;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytotoxicity associated immunodominant antigen (120 kDa protein).
GN CAGA OR CAT.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CCUG 17874 / NCTC 11638;
RX MEDLINE=93296225; Pubmed=8516329;
RA Covacci A., Censini S., Bugnoli M., Petraccia R., Burroni D.,
RA Macchia G., Massone A., Papini E., Xiang Z., Figura N.,
RA Rappuoli R.;
RT "Molecular characterization of the 128-kDa immunodominant antigen of
RT Helicobacter pylori associated with cytotoxicity and duodenal
RT ulcer."
RL Proc. Natl. Acad. Sci. U.S.A. 90:5791-5795(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CCUG 17874 / NCTC 11638;

```


SEQUENCE 1956 AA; 230285 MW; AE708AAE9009335 CRC64;

Query Match 6.8%; Score 130; DB 1; Length 1956;
 Best Local Similarity 19.2%; Pred. No. 0.68;
 Matches 90; Conservative 54; Mismatches 150; Indels 174; Gaps 19;

1 YE-KEQSLSD-----NKFVNNINSEYKQYETKATNDTFLNLSGKYCK 53
 1273 YECKNLVHIMKKKKKIKNNINNSNLY-----YHNITDTPKRNKREYCF 1323
 54 GGLPEKEDIFETNSADKGF-----YREXCOYCPDGVKDCGDKYTHKSDND 102
 1324 NKL-----LYKIOQKLLNHLNHLKXKKKNNYNDI-----DEVHLGNNNN 1367
 103 RERVNEDYRPMGVKPEPTNITVLYSGNEGD-----ITOKLEFNCN 143
 1368 NKNNSKREKPLKKNKKH-----RKNESNDNTFNTYSNNIHLKSKYVHHKNYYPD 1423
 144 SSTNYADKN-----OKWCYKQDENINCKLEONTNINNDPKT-----ISF 186
 1424 SCNLRKKNLSLEYNLKKYIYERKYLQCLKHDKYKVELEPRIDINYSYOMESIKT 1483
 187 HNFEE-----LWVTLR--DTIKMNDK--KTCINNTTTHCIDECNRCLCF 230
 1484 RNFHLSSEQPAFNSNLSFYIIKNDNNVYNNKNTYNNKNTKNS-----ICKNTYCN 1539
 231 DRWVOKKEEWN-----SIRKLF-----TKKNTI 254
 1540 KNYIYNNNNYNNKNNIYNNKNNILTHAKSVLLSGSKKFKFSNIIIRHKLKKNKKNI 1599
 255 QOQSYNNINLFFGYFPKVA-----DKIDKBAKKELEMEIKR----- 293
 1600 KRYKMNHNVTSGHIIILNCTHGFKKDYSSLNKRYIRVNNKRYMLNDVYDHHMYLT 1659
 294 -----KKNFSNLENNRDYLENAIELLDHLEKTAATCKDN 329
 1660 DWRCGTQYGGSKKKNNKNNNNNNNNILKKNINRLEHL--LVDKCKRN 1705

RESULT 7
 USOL YEAST STANDARD; PRT; 1790 AA.

ID USOL YEAST STANDARD; PRT; 1790 AA.

AC P25386;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Intracellular protein transport protein USOL.
 EN USOL OR INT1 OR YD058W.
 CC Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 CC NCBI_TaxID=4932;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=91180-1A;
 CC MEDLINE=91185402; PubMed=2010462;
 CC Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
 CC Yamaoka M.;
 CC "A cytoskeleton-related gene, usol, is required for intracellular
 CC protein transport in Saccharomyces cerevisiae.";
 CC J. Cell Biol. 113:245-260(1991).
 CC [2]
 CC SEQUENCE OF 782-1790 FROM N.A.
 CC Hostetler M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
 CC Kendrick K.E.;
 CC Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 CC [3]
 CC SEQUENCE OF 1-8 FROM N.A.
 CC Bai Y., Symington L.S.;
 CC Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
 CC COMPLEX.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR

MEMBRANS. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
 ER AND THE GOLGI COMPLEX.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
 CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
 CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
 CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YB047C FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X54378; CA38253.1; -
 CC EMBL; L03188; AAB00143.1; -
 CC EMBL; U53668; AAB66659.1; -
 CC PTR; A38455; A38455.
 CC SGD; S0002216; USOL.
 CC DR InterPro; IPR002017; Spectrin.
 CC DR Transport; Protein transport; Golgi stack; Cytoskeleton; coiled coil.
 CC KM DOMAIN 1 724 GLOBULAR HEAD.
 CC FT DOMAIN 1 725 1790 COILED COIL (POTENTIAL).
 CC FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
 CC FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
 CC FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
 CC FT CONFLICT 847 847 G -> E (IN REF. 2).
 CC FT CONFLICT 924 924 E -> K (IN REF. 2).
 CC FT CONFLICT 1253 1253 V -> I (IN REF. 2).
 CC FT CONFLICT 1319 1319 I -> V (IN REF. 2).
 CC FT CONFLICT 1461 1461 N -> S (IN REF. 2).
 CC FT CONFLICT 1581 1581 G -> S (IN REF. 2).
 CC FT CONFLICT 1600 1600 I -> V (IN REF. 2).
 CC FT CONFLICT 1661 1661 R -> S (IN REF. 2).
 CC FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
 CC FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
 CC SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Query Match 6.8%; Score 127.5; DB 1; Length 1790;
 Best Local Similarity 19.7%; Pred. No. 0.89;
 Matches 72; Conservative 63; Mismatches 140; Indels 91; Gaps 14;

9 LSNDKFNVINNSERYQFYEKLETOYATNDTFLNLSGKYCKGGLPE--KDIITF 65
 767 LTNEHKELD-----EKYQILNSSHSSLSKENFSILETEIKNVRDSIDEXTQLRDVLET 818
 66 NSADD-----KGIFRSEYCOYCPDGVKDCGDKYTHKSDNREBVNNDYRPMGVK 118
 819 KDKENOTALLEYSTTHKE-----DSIKTLENG--LETTLSOKKRAEDGIN 863
 119 PTNITVLYSGNEQDITOKLEFNCNSSTNYKDKNNQWCEYK-----D 162
 864 KMGKDLFALSREM-----QAVENCKNLOKEDKKNVNHQKETSLSKDAIKATEIKAIN 919
 163 ENINRCKLEONT---ETNDNPKIISFHNPELVYLLRDT--KNNDKIKTCINNTT 217
 920 ENLEEMKIQCNNSLSKEHEHLSKELYEYKSRQ-----SHDNLAKTEKLSKSIANN--- 970
 218 HCIDECNRNLCFPRWVOKKEEENSISIKLFTKKNNIQOQSYNNINLFFGY-----FF 271
 971 -----YKDMQANESLSIKAVEESKNSSISQLSMLNKKITDSMSQEKENFQ 1014
 272 KVMKDLKDEAKWKELEMEIKRKNKE--FSNLENNRDYLENAIELLDHLEKTAATCKDN 330
 1015 IERSISIKENIQKLTISDLEQTKKEIISDSSKDEYEQISLTKKL-ETATANDEN 1073
 331 TNEACE 336
 1074 VNKISE 1079

RESULT 8
 HMW2_MTCGE

ID HMW2_MYCCE STANDARD; PRT; 1805 AA.
AC P47460;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytadherence high molecular weight protein 2 (Cytadherence accessory
protein 2).
GN HMW2 OR MG218.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Nguyen D.T., Uterback R.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RA "The minimal gene complement of Mycoplasma genitalium.";
RA Science 270:397-403(1995).
[2]
RP SEQUENCE OF 557-659 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III,
RA "A survey of the Mycoplasma genitalium genome by using random
sequencing.";
RT J. Bacteriol. 175:7918-7930(1993).
RT -1- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHERIN PROTEINS
IN THE MYCOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
SIMILARITY).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL: U39701; AAC71437.1; -
CC EMBL: U02165; AAD12447.1; -
CC TIGR: MG218; -
CC Cytadherence; Structural protein; Coiled coil; Complete proteome.
FT DOMAIN 28 838 COILED COIL (POTENTIAL).
FT DOMAIN 914 1591 COILED COIL (POTENTIAL).
FT DOMAIN 1632 1723 COILED COIL (POTENTIAL).
FT DOMAIN 1777 1804 COILED COIL (POTENTIAL).
SQ SEQUENCE 1805 AA; 216252 MW; 11D093AF173284FD CRC64;
Query Match 6.4%; Score 122; DB 1; Length 1805;
Best Local Similarity 20.5%; Pred. No. 2;
Matches 81; Conservative 57; Mismatches 121; Indels 136; Gaps 18;
OY 6 OYSLNDNKKFVNINSEYKOFYEKLETOYATNTJFNLNKGKYGCGKPGEDIDFT 65
DB 1234 QTYLANKRK-----AEY-----SQOOLQKYTNLD-----LKENLERTD----- 1269
OY 66 NSADKKGIFYSEYQV---CPDGVKCDGKIKYTHKSDNDEVRNVEDYKPPWGVKPTNI 122
DB 1270 -OLKRC---HHSIFARKLRFKPNLFEKKOLKAKRIYVDKRNRLKENER----- 1315
OY 123 TVLSGNGOGITOKLEFNCSTNYKDKNNK-----WECYKDEINCKLE-- 171
DB 1316 NMLHLSNE-----TERKRAVLDEQISYFEKQKQATDALIASHKRKEGSLKILVLE 1371
OY 172 -QNTFINNDNPKIISFNFFELWYTLIDRTIKWMDKLK-----TCINNTTTHC 219

DB 1372 TRTKTKLNDFAFKSRQREFE-----NORLLLEQLTQOTSNNEFKTKA 1418
OY 220 IDECNENCICDFRWYKQKEEMNSIKLPTKKRNIOQSYNNINLFEQYFVKMDLKD 279
DB 1419 IOEINSS-----YKRGMEILNFOKKEFDKNN-----SLRYE-YFKMDEIER 1460
OY 280 DEAKKEKLEMNIRKKNES-----NLENNR-DYLENAIELLDLH----- 319
DB 1461 KESQVAVLEKTOFRKANLLEAOANKLITERKIDREKELKAFKDVDDIDSTNKORKE 1520
OY 320 -----KETATICKDNNTNACEF 337
DB 1521 LNEILNENKLLQOSSLIERERAINSKDSLANKIET 1555
RESULT 9
ID PVDB_PLAKN STANDARD; PRT; 1153 AA.
AC P50493;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Duffy receptor, beta form precursor (Erythrocyte binding protein).
OS Plasmodium knowlesi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5850;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92357776; PubMed=1496004;
RA Adams J.H., Sim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.,
RA "A family of erythrocyte binding proteins of malaria parasites.";
RA Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).
RT -1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
DETERMINANT.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: HIGH, TO P.VIVAX DUFFY RECEPTOR.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL: M90694; AAA29603.1; -
CC DR Malaria; Receptor; Glycoprotein; Signal; Transmembrane;
KW MultiGene family.
FT SIGNAL 1 21
FT CHAIN 22 1153
FT DOMAIN 22 1085
FT TRANSMEM 1086 1106
FT DOMAIN 1107 1153
FT CARBOHYD 134 134
FT CARBOHYD 179 179
FT CARBOHYD 199 199
FT CARBOHYD 202 202
FT CARBOHYD 252 252
FT CARBOHYD 348 348
FT CARBOHYD 430 430
FT CARBOHYD 467 467
FT CARBOHYD 576 576
FT CARBOHYD 626 626
FT CARBOHYD 722 722
FT CARBOHYD 847 847
FT CARBOHYD 856 856
FT CARBOHYD 900 900
FT CARBOHYD 910 910
FT CARBOHYD 935 935
SQ SEQUENCE 1153 AA; 130471 MW; 6497BD1GCEVBE01 CRC64;
Query Match 6.3%; Score 121; DB 1; Length 1153;
Best Local Similarity 19.5%; Pred. No. 14;


```

RESULT 10
ARP_PLAFA
ID ARP_PLAFA STANDARD; PRT; 537 AA.
AC P04931;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Asparagine-rich protein (AG319) (ARP) (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI TaxID=5833;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=66176787; Pubmed=2421257;
RA Stahl H.-D., Bianco A.E., Crewther P.E., Burkot T., Coppel R.L.,
RB Brown G.V., Anders R.F., Kemp D.J.;
RT "An asparagine-rich protein from blood stages of Plasmodium
RL falciparum shares determinants with sporozoites." ;
RC Nucleic Acids Res. 14:3089-3102(1986).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb.ch/announce/
or send an email to license@isb.slb.ch).
-----
CC
DR EMBL; M24328; AAA29491.1; -.
DR PIR; A23770; A23770.
KW Malaria.
FT NON TER      1          1
FT FT           537       537
SQ SEQUENCE    537 AA; 63000 MW; AD6D76B15318CC239 CRC64;
Query Match              6.3%; Score 120.5; DB 1; Length 537;
Best Local Similarity   19.3%; Pred. No. 0.65;
Matches 76; Conservative 73; Mismatches 156; Indels 89; Gaps 20;

OY 5 IOSISYNDRKEPVNNINSEYYKOFIEKLKETQATMDTFNLINLECK--YCKGGLPEEKD 61
Db 60 VEEHRRNNSIDNNRNNINNNTN---OOTRRS---SFEMENENENKNKYHHGM--NNN 109
OY 62 IFFTSADDKGFIFYSESYCYQCVPDCGVACD---GIKTTHKSNDNR----ERYNN-EDYKP 113
DB 110 IHFKKKYIDNNSSMKM-----TDNKTKTDSIYNMGKITNNDDNMMDYLRIANNINNEYK- 161
```

```

RESULT 11
Y091_CAEEL
ID Y091_CAEEL STANDARD; PRT; 782 AA.
AC Q10030;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical 90.1 kDa protein C27D6.1 in chromosome II.
GN C27D6.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
OC Rhabditiidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol NZ;
RL Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; U23179; AAC64721.1; -.
DR WormPep; C27D6.1; CE01811.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 669 672 POLY-ALA.
FT DOMAIN 223 331 COILED COIL (POTENTIAL).
FT DOMAIN 348 398 COILED COIL (POTENTIAL).
FT DOMAIN 428 601 COILED COIL (POTENTIAL).
FT DOMAIN 699 743 COILED COIL (POTENTIAL).
SQ SEQUENCE 782 AA; 90059 MW; BB7CB842CD7273BB CRC64;

Query Match 6.2%; Score 119.5; DB 1; Length 782;
Best Local Similarity 21.1%; Pred. No. 1.2; Mismatches 78; Indels 69; Gaps 9;
Matches 50; Conservative 40;

160 YK---DENINCKLEQONTNEINNNDNPKRITSFHNFEFLMTWYLLRDT----- 201
Db 181 YKRAGDAASHMSKRDLSPPSVSFTFTVLAVFHFKVPVAATYITIREKRSERQTKLESLE 240
Qy 202 -IKWMDKLKTCINTNTTHCIDECNEN-----CLCEPDW-VKOE----- 238
Db 241 RLKANERARKETIEAERAKWKDRATRNSKRPLPELELAETVOAKEMQVGSGEIMQNKQ 300
Qy 239 --EEENSIAKLTFTKKKNINQOISTYSNI---NNLFESYEFKV-----MDKLKDKE 281

```


Db 301 LVEELNEVOKKLEIENSOKTFHOKVSTLNDIEEFQNPDEQDQSLQFMVLEQEI 360
 QY 282 AKKMELENIKKR-----KNFSNLNNRNYLNALELDLH-----KEATATCK 327
 Db 361 VAYQKCIKQENELKELKLOQLSSSLTVNHNHSTLMDHLEINKESQRETOGICK 417

RESULT 12
 Y228_BORBU STANDARD; PRT; 971 AA.
 ID Y228_BORBU
 AC 051246;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein BB0228.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 GN Bb0228.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 SEQUENCE FROM N.A.
 STRAIN=ATCC 35210 / B31;
 MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Kellavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer M., Adams M.D., Gocayne J.D., Weidman J.,
 RA Utermack T., Wathley L., McDonald L., Artiach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia
 burgdorferi.";
 RL Nature 380:580-586(1997).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE001133; AAC66621.1;
 DR TIGR; BB0228;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 971 AA; 112959 MW; 088A68BD7B8C591A CRC64;

Query Match
 Best Local Similarity 23.1%; Score 119; DB 1; Length 971;
 Matches 82; Conservative 42; Mismatches 121; Indels 110; Gaps 19;

QY 4 ELDSYLSNDKRFVNNINSEYKOFYEKLEKTOYATNDTFLNINLNEG-KYCGGLDGE-K 60
 Db 156 EMGGSYSNNKSLNLEIVSS-----SLFEGAVYYDSGGGIPNTII 194
 QY 61 DITFTNSADDKGIFRSEYKQVCPDGVKCDGIKTKHSKSDNRERNNDYKPPMGVKPT 120
 Db 195 DLYVESFLD-----FYKKY--TLENCKIFLQGNTOLEKMLNTEKTIIRPYKE----- 242
 QY 121 NITVYLSNGEGDITQKLENFCSNTNYKDKNNQWE-----CYR--DENINR----- 167
 Db 243 -----KSNVNININIEVKKRWEKGRKLTJYKIKPKENDSLGVYIN 280
 QY 168 --CKLEQNTENNDPKIISFNFELWYTLRDITKMN-DKLTGCIANNTHHCIDECN 224
 Db 281 WIC-----TEINNIIDST-----GEILSEILLDSCSTTINILKSGIEDIAH-ISGIN 329
 QY 225 RN-----CLCPDRVKKOKKEEEMSIKRLFTKKKNIQOSYYSN--INMLFEGYFPKVDK 276
 Db 330 TDLKESIFSPGLONVVENKEKEFKML--VFSPLKVLVKNKIKPKELIKGLGYEPAKKE 387
 QY 277 -----LDKDEAKWMELENIKKR-----NEFSN-LENNRNYLENAIELL 316

Db 388 KGQNFPIALMIKSPKGLNGLHPKIKTQTSYIINEITNKEKIYFENLEIYTL 442

RESULT 13
 PYDA_PLAKN STANDARD; PRT; 1073 AA.
 ID PYDA_PLAKN
 AC P22345;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Duffy receptor, alpha form precursor (Erythrocyte binding protein).
 OS Plasmodium knowlesi.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5850;
 RN [1]
 SEQUENCE FROM N.A.
 MEDLINE=92357776; PubMed=1496004;
 RA Adams J.H., Sim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.,
 RA "A family of erythrocyte binding proteins of malaria parasites.";
 RA Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).
 RN [2]
 SEQUENCE OF 296-1073 FROM N.A.
 MEDLINE=91004213; PubMed=2170017;
 RA Adams J.H., Hudson D.E., Torii M., Ward G.E., Wellens T.E.,
 RA Aikawa M., Miller L.H.;
 RT "The Duffy receptor family of Plasmodium knowlesi is located within
 RT the micronemes of invasive malaria merozoites.";
 RL Cell 63:141-153(1990).

CC -1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
 CC DETERMINANT.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: HIGH TO P.VIVAX DUFFY RECEPTOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M90466; AAA29602.1;
 DR EMBL; M68517; AAA29590.1;
 DR EMBL; M68518; AAA29591.1;
 DR PIR; A35970; A35970.
 KW Malaria; Receptor; Glycoprotein; Signal; Transmembrane;
 KM Multigene family.
 KW SIGNAL

FT CHAIN 1 21
 FT SIGNAL 1 21
 FT DOMAIN 22 1073
 FT TRANSSEM 1008 1029
 FT DOMAIN 1030 1073
 FT CARBOHYD 134 134
 FT CARBOHYD 179 179
 FT CARBOHYD 202 179
 FT CARBOHYD 202 202
 FT CARBOHYD 252 252
 FT CARBOHYD 348 348
 FT CARBOHYD 679 679
 FT CARBOHYD 746 746
 FT CARBOHYD 779 779
 FT CARBOHYD 788 788
 SQ SEQUENCE 1073 AA; 120683 MW; 3965FC9F46B71808 CRC64;

Query Match
 Best Local Similarity 20.0%; Score 119; DB 1; Length 1073;
 Matches 81; Conservative 64; Mismatches 149; Indels 110; Gaps 18;

QY 7 SYLSNDKRFVNNINSEYKOFYEKLEKTOYATNDTFL--NLNE--GYCKG-----GLPG 58
 Db 175 NFELENSK-----GDPHPYNNRKRERNSGVINOTFLONNVVNDKCDKRRGRMDGCA 229
 QY 59 EKDI-----FTNSADDKGIFRSEYKQVCPDGVKCDGIKTKHSKSDNRERNNDYKPPMGVKPT 120

```

Db 230 EXDICSVRRYOLCMKGLTNVNTPTSHNDITFLKLNKRLKMLDAVAEGDILL- 285
OY 99 SMDREYVNEDEKPP-----MGVKTPTNITVYSGNEGDIDIOKLEFNCNS--STNYKD 150
Db 286 -----KKNYQYNKECKKIRNGLDGFIDITIMKTNEGIGYSOVANNRQYFVIGIDEKA 339
OY 151 KNNOK-----WECY-----YKDNINRCKLEONTEINNDPKIISPHNEFLW 193
Db 340 KORKOWMNSKEHIMRAMFIRSRLKEKFWIKCKDVLTKY---EPQI---YRWIREW 393
OY 194 -----VYLLADDTIKANDKL--KTCINNTTTCIDECNRNCLCPRWYKQKEEENSTIKL 247
Db 394 GRDYMSKLPKEQGLNKCKASKLYNNMAICMPLCHDCKSYDOWITRRKKQMDVLSTK 453
OY 248 FTKKKNIOQSYNINNLFGYFEFFKVMKLDKDEAKKKELEMINIKRKNESNLNNRDY 307
Db 454 FSSYKTKQKGTENIATAY-----DILKQELNGFEKA-----TFENIENKRDN 496
308 LENAIELLDHLKETATICKDNNTNEACET-----SHNATNP 345
497 LYN-----HLCPVVEARKNTQENVKVNGSGVESKASSNP 533

```

```

RESULT 14
IFH1_YEAST STANDARD; PRT; 1085 AA.
AC P39520;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE IFH1 protein (RRP3 protein).
GN IFH1 OR RRP3 OR YLR223C OR L8083.9.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FL100;
RX MEDLINE=95304839; PubMed=7785326;
RA Chereil I., Thuriaux P.;
RT "The IFH1 gene product interacts with a fork head protein in
RL Saccharomyces cerevisiae."
RT yeast 11:261-270(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=5288c / AB972;
RX Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favallo A., Fulton L., Gatlung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifken L., Riles L., Taich A., Trevasakis E., Vignati D.,
RA Wilcox L., Wohlman P., Yaudin M., Wilson R., Waterston R.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: CONTROLS THE PRE-RNA PROCESSING MACHINERY IN
CC CONJUNCTION WITH FHL1. COULD CONVERT FHL1 FROM A REPRESSOR
CC TO AN ACTIVATOR.
CC - SUBCELLULAR LOCATION: Nuclear (Probable).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z29488; CA82624.1;
CC EMBL: U19027; AAB67412.1;
CC PIR: S47477; S47477.
CC SGD: S0004213; IFH1.

```

```

KW Nuclear protein; Transcription regulation.
FT DOMAIN 122 163 ASP/GLU-RICH (HIGHLY ACIDIC).
SQ SEQUENCE 1085 AA; 122491 MW; BEIC/DEF06213FE0 CRC64;
Query Match 6.2%; Score 118.5; DB 1; Length 1085;
Best Local Similarity 20.3%; Pred. No. 2;
Matches 86; Conservative 66; Mismatches 153; Indels 119; Gaps 15;
OY 10 SMDKFFV-----NNINSEYKQFYERKLEKQYATNDTFLNLEBKGYCKGIPGKIDTFT 65
Db 199 SNKKSYSKLPKKNENEDEQKEKEKEKEDEQKQESNKKTEVNG--SGTTTQOALSPK 254
OY 66 NSADDKGIFTRSEYQYCPDQYKC-----DGIYTHKSD 100
Db 255 FKEDDEISFGNGEGYNEIDGEEVLDLKKNNNGNEEDKLDKSYMLGNDLFRPINS 314
OY 101 NDRER-----VNNEDYKRPWGKPTNITVLVSG-----NEGDIOKLEEN- 140
Db 315 SDESEYIDODAYFDVYINNEDSHGEIG-----TDLEFGEDDLPLDEEEDNIVSEIQND 368
OY 141 -----FCNS-----STNYKDKNNQKWEKYKDNINRCKLEONTEINND-----NPKTI 184
Db 369 DELSFDSIHEGSDPYEDAKNFKLQNEVNGENGYDEEDDEDEIDMSDFMPYEDPKFA 428
OY 185 SFHNFELWYTYLLRDITKNDKLTCTINNTTTCIDECNRNCLCFDRWYKQKEEENSTI 244
Db 429 NLXYFGD-----GSEPKLSLSTPLMLNDELKSLKREAKKREOR 475
OY 245 KLPFTKKKNIOQSYNINNLFGYFEFFKVMKLDKDEAKKKELEMINIKRKNESNLNNRDY 298
Db 476 KLTYYKQKQESTRTTSVDN--DEYINVFQSDENSGHKSCKGRHKSGLSHENKKG 533
OY 299 SNLENNRDYLENAI-----ELLDLHKLKETATICKDNNTNEACETSHN 340
Db 534 SNLKSDNDELPSRHSYVNLNGSKYDSDSDYDNLILD---VAMHPSDESESES-ETSHD 588
OY 341 ATTN 344
Db 589 ADTD 592
RESULT 15
IF2P_HUMAN STANDARD; PRT; 1220 AA.
AC O60841; O95805; Q9UFB1; Q9UMN7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Translation Initiation Factor IF-2.
GN IF2 OR KIA0741.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RL for large proteins in vitro."
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=99362399; PubMed=10433305;
RA Wilson S.A., Sleiro-Vazquez C., Edwards N.J., Iourin O., Byles E.D.,
RA Kotsopoulos E., Adamson C.S., Kingsman S.M., Kingsman A.J.,
RA Martin-Rendon E.;
RT "Cloning and characterization of hIF2, a human homologue of bacterial
RT translation initiation factor 2 and its interaction with HIV-1

```

RT matrix.";
 RL Biochem. J. 342:97-103(1999).
 RN [3]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=Testis;
 RX MEDLINE=99218282; PubMed=10200264;
 RA Lee J.H., Choi S.K., Roll-Mecak A., Burley S.K., Dever T.E.;
 RT "Universal conservation in translation initiation revealed by human
 RT and archaeal homologs of bacterial translation initiation factor
 RT IF2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:4342-4347(1999).
 RN [4]
 RP SEQUENCE OF 89-1220 FROM N.A.
 RC TISSUE=Testis;
 RA Koerner K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
 RL Submitted (Dec-1999) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 833-1220 FROM N.A.
 RX MEDLINE=21064499; PubMed=11124703;
 RA Stanchi F., Bertocco E., Toppo S., Dioguardi R., Simionati B.,
 RA Camata N., Zimbello R., Ianfranchi G., Valle G.;
 RT "Characterization of 16 novel human genes showing high similarity to
 RT yeast sequences.";
 RL Yeast 18:69-80(2001).
 CC -1- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING
 CC THE BINDING OF THE FORMYL METHIONINE-tRNA TO RIBOSOMES. SEEKS TO
 CC FUNCTION ALONG WITH EIF-2 (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AB018284; BAA34461.1; -;
 DR EMBL: AJ006776; CAB44357.1; -;
 DR EMBL: AF078035; AAD16006.1; -;
 DR EMBL: AL133563; CAB63717.1; -;
 DR EMBL: AJ006412; CA07018.1; -;
 DR MIM: 606086; -;
 DR InterPro: IPR004161; EFTU_D2.
 DR InterPro: IPR00795; EF_GTPbind.
 DR InterPro: IPR00178; IF2.
 DR InterPro: IPR005225; Small_GTP.
 DR Pfam: PF00009; GTP_EFTU; 1.
 DR Pfam: PF03144; GTP_EFTU_D2; 2.
 DR PRINTS: PR00315; ELONGATNCT.
 DR TIGRFAWS: TIGR00231; small_GTP; 1.
 DR PROSITE: PS01176; IF2; FALSE_NEG.
 KW Initiation factor; protein biosynthesis; GTP-binding.
 FT NP_BIND 638 645
 FT DOMAIN 39 50
 FT POLY-LYS 94 99
 FT POLY-LYS 138 142
 FT DOMAIN 313 322
 FT POLY-ASP 353 356
 FT DOMAIN 361 364
 FT POLY-GLU 491 496
 FT POLY-GLU 529 567
 FT MUTAGEN 640 640
 FT MUTAGEN 706 706
 FT MUTAGEN 706 706
 FT MUTAGEN 759 759
 FT CONFLICT 64 64
 FT CONFLICT 92 92
 FT CONFLICT 180 180
 FT I -> M (IN REF. 3).
 FT E -> G (IN REF. 1).
 FT T -> I (IN REF. 2).
 FT I -> M (IN REF. 3).

FT CONFLICT 256 256 K -> R (IN REF. 3).
 FT CONFLICT 522 522 T -> K (IN REF. 1).
 FT CONFLICT 549 549 E -> V (IN REF. 3).
 FT CONFLICT 669 669 G -> W (IN REF. 3).
 FT CONFLICT 894 894 E -> K (IN REF. 2).
 SO SEQUENCE 1220 AA; 138799 MW; 010CC15DE475E5B8 CRC64;
 Query Match 6.2%; Score 118.5; DB 1; Length 1220;
 Best Local Similarity 21.2%; Pred. No. 2.2;
 Matches 68; Conservative 53; Mismatches 132; Indels 67; Gaps 11;
 QY 29 EKLEKQVATNPTFLNLNEGKCKGGLGPKGKIDITFTNSADKGIYREYQVCPDGV 88
 DB 45 KKKKKQDFDEDDILKELELESLDAOGIKADRETVAVKPTENNEEETSK-----DK 96
 QY 89 KCDGIKVTYHKS--DNDREHVNNDYKPPWGVKPTNITVLVSGNEGIDITOKLENFCNSST 146
 DB 97 KKKGGKQKQSFDDNDSELEEDKDSKKTATPK--VEYSSGSDDDDDNKLP---KKAK 151
 QY 147 NTKDKNNQWECYKADENINRCKLEONTIENNNDPKIISFHFELWVYLLRDTIKWD 206
 DB 152 GRAQSKNKKWDSEDED--NSKKIKERSRINSGESGDSDEFLO-----SRKGQKKQ 204
 QY 207 KKTCTINNTTTCIDECNKNCLCFDRWVKQKEENNSIKKLTKKKNIGQSYYSNNLF 266
 DB 205 KKKPGNISGMDDDASFT---KTVAKKAE---KKERERRK----- 243
 QY 267 EGYFFKMDLKDDEAKKELMENIK---RRKNFSLNNDYLENAIELL----- 316
 DB 244 -----DEKAKLKLKKELELETKGKQDSQKQKESQKFEETVYKSKVYDTGVI 292
 QY 317 ---DHKETATICKDNNTNE 333
 DB 293 PASEEKAEPTAAEDDNEGD 312

Search completed: April 28, 2003, 10:30:19
 Job time : 13.3149 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 12.9282 Seconds

(without alignments)
5498.552 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747

Sequence: 1 YKEKIQSYLSNDNKFVNININ.....CRDNTNACEFSHATTNP 345

Scoring table: BL0SUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

al number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP mhc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *
15: SP virus: *
16: SP bacteriophage: *
17: SP archaea: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1918	100.0	3542	5	Q9U5M2
2	1864	97.2	510	5	Q8T6L0
3	1766	92.1	492	5	Q8T6K9
4	1765	92.0	494	5	Q8T6K7
5	1718	89.6	1615	5	Q8T325
6	1717	89.5	3287	5	Q8T325
7	1631	85.0	465	5	Q8T6K4
8	1625	84.7	461	5	Q8T6K3
9	1610	83.9	492	5	Q8T6K6
10	1609	83.9	494	5	Q8T6K0
11	1603	83.6	496	5	Q8T6K8
12	1577	82.2	494	5	Q8T6K5
13	1465	76.4	460	5	Q8T6K1
14	1260	65.7	427	5	Q8T6K2
15	771	40.2	3006	5	Q26032
16	513	26.7	1665	5	Q9U4A2

17	501.5	26.1	2647	5	P90580	P90580	plasmodium
18	499.5	26.0	2212	5	Q94657	Q94657	plasmodium
19	457	23.8	2664	5	Q26033	Q26033	plasmodium
20	451	23.5	2169	5	Q97312	Q97312	plasmodium
21	446	23.3	2209	5	Q9U0G6	Q9U0G6	plasmodium
22	446	23.3	2228	5	Q60991	Q60991	plasmodium
23	444	23.1	2270	5	Q9XZB8	Q9XZB8	plasmodium
24	442.5	23.1	2527	5	Q95W83	Q95W83	plasmodium
25	437	22.8	3026	5	Q26030	Q26030	plasmodium
26	432.5	22.5	2197	5	Q96296	Q96296	plasmodium
27	429	22.4	2277	5	Q9U0G5	Q9U0G5	plasmodium
28	407	22.2	480	5	Q9NAV6	Q9NAV6	plasmodium
29	403.5	21.0	2706	5	Q15870	Q15870	plasmodium
30	399.5	20.8	2658	5	Q8T5G0	Q8T5G0	plasmodium
31	394.5	20.6	3078	5	Q26031	Q26031	plasmodium
32	380	19.8	2182	5	Q26034	Q26034	plasmodium
33	363	18.9	1711	5	Q96108	Q96108	plasmodium
34	362	18.9	2209	5	Q97324	Q97324	plasmodium
35	361.5	18.8	2135	5	Q61077	Q61077	plasmodium
36	357.5	18.6	1729	5	Q25734	Q25734	plasmodium
37	357.5	18.6	2924	5	Q25733	Q25733	plasmodium
38	356	18.6	2163	5	Q9NFB6	Q9NFB6	plasmodium
39	331	17.3	2042	5	Q25766	Q25766	plasmodium
40	231	12.0	921	5	Q25989	Q25989	plasmodium
41	223.5	11.7	174	5	Q15664	Q15664	plasmodium
42	223.5	11.7	174	5	Q18671	Q18671	plasmodium
43	219	11.4	130	5	Q15662	Q15662	plasmodium
44	216.5	11.3	174	5	Q15652	Q15652	plasmodium
45	216.5	11.3	174	5	Q15665	Q15665	plasmodium

ALIGNMENTS

RESULT 1

Q9U5M2 PRELIMINARY; PRT; 3542 AA.
AC Q9U5M2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE PCR3 CSA ligand (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCR3;
RX MEDLINE=20006305; PubMed=10535993;
RA Buffet P., Gamain B., Scheidig C., Baruch B., Oishi S., Fujii N.,
RA Fusai T., Parzy D., Miller L.H., Gysin J., Scherf A.;
RT "Plasmodium falciparum domain mediating adhesion to chondroitin
sulfate A: A receptor for human placental infection.";
RL EMBL; AJ13811; CAB59840.1; -
DR InterPro: IPR001219; Neurotoxin.
DR InterPro: IPR004258; PFEEMP.
DR Pfam: PF03011; PFEEMP.1.
DR PRINTS; PR00284; TOXIN.
FT NON_TER 3542 3542
SQ SEQUENCE 3542 AA; 413089 MW; 970D85EE88DA2EC2 CRC64;

Query Match 100.0%; Score 1918; DB 5; Length 3542;

Best Local Similarity 100.0%; Pred. No. 7.5e+108; Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YKEKIQSYLSNDNKFVNINSEYKQFEKLETOYATNDFLNLNKGKCKGGLPEK 60
DB 403 YKEKIQSYLSNDNKFVNINSEYKQFEKLETOYATNDFLNLNKGKCKGGLPEK 462
QY 61 DITFTNSADDKGIRYRSYCYCPCGCKGICIKTKHSNDREVRNNEDEKPPGKAPT 120
|||||

DB 463 DITFTNSADDKGIFYSSEYQVCPDGVKCDGKIKYTHKSDNDRERNNEDYKPPMGVCKPT 522
QY 121 NITVLYSGNEOGDITOKLENFCNSSTNYKDKNNQKWECCYKDEININCKLEQTEINNDN 180
DB 523 NITVLYSGNEOGDITOKLENFCNSSTNYKDKNNQKWECCYKDEININCKLEQTEINNDN 562
QY 181 PKIISHNFEFLWVYLLRPTIKWMDKLTCTCINNNTTHCIDEENRNCCLCFDRWVKOKEE 240
DB 583 PKIISHNFEFLWVYLLRPTIKWMDKLTCTCINNNTTHCIDEENRNCCLCFDRWVKOKEE 642
QY 241 WNSIKKLFTRKKNNIOOSYSYNNINNFEGYFFKVMKLDKDEAKWELMENIKRKKNEFSN 300
DB 643 WNSIKKLFTRKKNNIOOSYSYNNINNFEGYFFKVMKLDKDEAKWELMENIKRKKNEFSN 702
QY 301 LENNRDYLENAIEILLDHLKETATICKDNTNACETSHNATNP 345
DB 703 LENNRDYLENAIEILLDHLKETATICKDNTNACETSHNATNP 747

RESULT 2

Q8T6L0 PRELIMINARY; PRT; 510 AA.
Q8T6L0;
AC 08T6L0;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
RT implicated in malaria in pregnancy."
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL: AF458886; AAL84269.1; -.
FT NON-TER 1
FT NON-TER 510
SQ SEQUENCE 510 AA; 60713 MW; E3CF55DC04601099 CRC64;

Query Match 97.2%; Score 1864; DB 5; Length 510;
Best Local Similarity 98.5%; Pred. No. 1.9e-105;
Matches 335; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YEKELSYLSNDKNEFVNNINSEYKOFYKELKETOYATNDTFLINLEGGYCKGGLPGER 60
DB 171 YEKELSYLSNDKNEFVNNINSEYKOFYKELKETOYATNDTFLINLEGGYCKGGLPGER 230
QY 61 DITFTNSADDKGIFYSSEYQVCPDGVKCDGKIKYTHKSDNDRERNNEDYKPPMGVCKPT 120
DB 231 DITFTNSADDKGIFYSSEYQVCPDGVKCDGKIKYTHKSDNDRERNNEDYKPPMGVCKPT 290
QY 121 NITVLYSGNEOGDITOKLENFCNSSTNYKDKNNQKWECCYKDEININCKLEQTEINNDN 180
DB 291 NITVLYSGNEOGDITOKLENFCNSSTNYKDKNNQKWECCYKDEININCKLEQTEINNDN 350
QY 181 PKIISHNFEFLWVYLLRPTIKWMDKLTCTCINNNTTHCIDEENRNCCLCFDRWVKOKEE 240
DB 351 PKIISHNFEFLWVYLLRPTIKWMDKLTCTCINNNTTHCIDEENRNCCLCFDRWVKOKEE 410
QY 241 WNSIKKLFTRKKNNIOOSYSYNNINNFEGYFFKVMKLDKDEAKWELMENIKRKKNEFSN 300
DB 411 WNSIKKLFTRKKNNIOOSYSYNNINNFEGYFFKVMKLDKDEAKWELMENIKRKKNEFSN 470
QY 301 LENNRDYLENAIEILLDHLKETATICKDNTNACETSHN 340
DB 471 LENNRDYLENAIEILLDHLKETATICKDNTNACETSHN 510

RESULT 3

Q8T6K9

ID 08T6K9 PRELIMINARY; PRT; 492 AA.
AC 08T6K9;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
RT implicated in malaria in pregnancy."
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL: AF458889; AAL84272.1; -.
FT NON-TER 1
FT NON-TER 492
SQ SEQUENCE 492 AA; 58590 MW; 277204FDF55014C6 CRC64;

Query Match 92.1%; Score 1766; DB 5; Length 492;
Best Local Similarity 97.5%; Pred. No. 1.6e-99;
Matches 316; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEKELSYLSNDKNEFVNNINSEYKOFYKELKETOYATNDTFLINLEGGYCKGGLPGER 60
DB 169 YEKELSYLSNDKNEFVNNINSEYKOFYKELKETOYATNDTFLINLEGGYCKGGLPGER 228
QY 61 DITFTNSADDKGIFYSSEYQVCPDGVKCDGKIKYTHKSDNDRERNNEDYKPPMGVCKPT 120
DB 229 DITFTNSADDKGIFYSSEYQVCPDGVKCDGKIKYTHKSDNDRERNNEDYKPPMGVCKPT 288
QY 121 NITVLYSGNEOGDITOKLENFCNSSTNYKDKNNQKWECCYKDEININCKLEQTEINNDN 180
DB 289 NITVLYSGNEOGDITOKLENFCNSSTNYKDKNNQKWECCYKDEININCKLEQTEINNDN 348
QY 181 PKIISHNFEFLWVYLLRPTIKWMDKLTCTCINNNTTHCIDEENRNCCLCFDRWVKOKEE 240
DB 349 PKIISHNFEFLWVYLLRPTIKWMDKLTCTCINNNTTHCIDEENRNCCLCFDRWVKOKEE 408
QY 241 WNSIKKLFTRKKNNIOOSYSYNNINNFEGYFFKVMKLDKDEAKWELMENIKRKKNEFSN 300
DB 409 WNSIKKLFTRKKNNIOOSYSYNNINNFEGYFFKVMKLDKDEAKWELMENIKRKKNEFSN 468
QY 301 LENNRDYLENAIEILLDHLKETAT 324
DB 469 LENNRDYLENAIEILLDHLKETAT 492

RESULT 4

Q8T6K7 PRELIMINARY; PRT; 494 AA.

ID 08T6K7
AC 08T6K7;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
RT implicated in malaria in pregnancy."
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL: AF458889; AAL84272.1; -.
FT NON-TER 1
FT NON-TER 494
SQ SEQUENCE 494 AA; 58773 MW; C6D6731157A1C13A CRC64;

Query Match 92.0%; Score 1765; DB 5; Length 494;
Best Local Similarity 96.6%; Pred. No. 1.9e-99;
Matches 315; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 1 YEKEIOSYLSNDKFFVNNINSEYKOFYEKLETOYATNDTFLNLNEGKCGGLPGEK 60
DB 169 YEKEIOSYLSNDKFFVNNINSEYKOFYEKLETOYATNDTFLNLNEGKCGGLPGEK 228
QY 61 DIFTNSADDKGIFRSEYCOVCPDCGVKCDGKITYTHKSDNDRVANNEDYKPPMGVKPT 120
DB 229 DIFTNSADDKGIFRSEYCOVCPDCGVKCDGKITYTHKSDNDRVANNEDYKPPMGVKPT 288
QY 121 NITVLVSGNEQGDITQKLENFCSSTNYKDKNNQKWECCYKEDENINRCKLEQNTTEINNDN 180
DB 289 NITVLVSGNEQGDITQKLENFCSSTNYKDKNNQKWECCYKEDENINRCKLEQNTTEINNDN 348
QY 181 PRTISFHNFEFLWVYLLRDTIKMNDKLTCTINNTHHCIDECNRNCLCFDRVWVKOKEE 240
DB 349 PRTISFHNFEFLWVYLLRDTIKMNDKLTCTINNTHHCIDECNRNCLCFDRVWVKOKEE 408
QY 241 WNSIKKLFTRKKNIQOOSYSNINNLEFEGYFFKVMKLDKDEAKWKLMEINIRKKNEFSN 300
DB 409 WNSIKKLFTRKKNNVPOPYTININNLEFEGYFFKVMKLDKDEAKWKLMEINIRKKNEFSN 468
QY 301 LENNRDYLENAIELLDHLKETATTC 326
DB 469 LENNRDYLENAIELLDHLKETATTC 494

RESULT 5
Q8T325 PRELIMINARY; PRT; 1615 AA.
AC Q8T325;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE PFEMP1 (Fragment).
GN TM284VAR3.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TM284;
RX MEDLINE-21927235; PubMed-11930336;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implicated in malaria in pregnancy."
J. Infect. Dis. 185:1207-1211(2002).
EMBL; AJ420412; CAD20868.1; -.
FT NON_TER 1615
SQ SEQUENCE 1615 AA; 188639 MW; FCA896C00B6DEA6 CRC64;

Query Match 89.6%; Score 1718; DB 5; Length 1615;
Best Local Similarity 89.3%; Pred. No. 4.3e-96;
Matches 308; Conservative 14; Mismatches 23; Indels 0; Gaps 0;

QY 1 YEKEIOSYLSNDKFFVNNINSEYKOFYEKLETOYATNDTFLNLNEGKCGGLPGEK 60
DB 393 YEKEIOSYLSNDKFFVNNINSEYKOFHDLRKNKNLDTFLNLNEGKCGGLPGEK 452
QY 61 DIFTNSADDKGIFRSEYCOVCPDCGVKCDGKITYTHKSDNDRVANNEDYKPPMGVKPT 120
DB 453 DIFTNSADDKGIFRSEYCOVCPDCGVKCDGKITYTHKSDNDRVANNEDYKPPMGVKPT 512
QY 121 NITVLVSGNEQGDITQKLENFCSSTNYKDKNNQKWECCYKEDENINRCKLEQNTTEINNDN 180
DB 513 NITVLVSGNEQGDITQKLENFCSSTNYKDKNNQKWECCYKEDENINRCKLEQNTTEINNDN 572
QY 181 PRTISFHNFEFLWVYLLRDTIKMNDKLTCTINNTHHCIDECNRNCLCFDRVWVKOKEE 240
DB 573 PRTISFHNFEFLWVYLLRDTIKMNDKLTCTINNTHHCIDECNRNCLCFDRVWVKOKEE 632

QY 241 WNSIKKLFTRKKNIQOOSYSNINNLEFEGYFFKVMKLDKDEAKWKLMEINIRKKNEFSN 300
DB 633 WNSIKKLFTRKKNNVPOPYTININNLEFEGYFFKVMKLDKDEAKWKLMEINIRKKNEFSN 692
QY 301 LENNRDYLENAIELLDHLKETATTCNDNTNACETSHNATNP 345
DB 693 LENNRDYLENAIELLDHLKETATTCNDNTNACETSHNATNP 737

RESULT 6
Q8T326 PRELIMINARY; PRT; 3287 AA.
AC Q8T326;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE PFEMP1 (Fragment).
GN TM180VAR2.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TM180;
RX MEDLINE-21927235; PubMed-11930336;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implicated in malaria in pregnancy."
J. Infect. Dis. 185:1207-1211(2002).
EMBL; AJ420411; CAD20867.1; -.
FT NON_TER 3287
SQ SEQUENCE 3287 AA; 383550 MW; 58F8F86FC244536 CRC64;

Query Match 89.5%; Score 1717; DB 5; Length 3287;
Best Local Similarity 89.0%; Pred. No. 1e-95;
Matches 307; Conservative 17; Mismatches 21; Indels 0; Gaps 0;

QY 1 YEKEIOSYLSNDKFFVNNINSEYKOFYEKLETOYATNDTFLNLNEGKCGGLPGEK 60
DB 393 YEKEIOSYLSNDKFFVNNINSEYKOFHDLRKNKNLDTFLNLNEGKCGGLPGEK 452
QY 61 DIFTNSADDKGIFRSEYCOVCPDCGVKCDGKITYTHKSDNDRVANNEDYKPPMGVKPT 120
DB 453 DIFTNSADDKGIFRSEYCOVCPDCGVKCDGKITYTHKSDNDRVANNEDYKPPMGVKPT 512
QY 121 NITVLVSGNEQGDITQKLENFCSSTNYKDKNNQKWECCYKEDENINRCKLEQNTTEINNDN 180
DB 513 NITVLVSGNEQGDITQKLENFCSSTNYKDKNNQKWECCYKEDENINRCKLEQNTTEINNDN 572
QY 181 PRTISFHNFEFLWVYLLRDTIKMNDKLTCTINNTHHCIDECNRNCLCFDRVWVKOKEE 240
DB 573 PRTISFHNFEFLWVYLLRDTIKMNDKLTCTINNTHHCIDECNRNCLCFDRVWVKOKEE 632
QY 241 WNSIKKLFTRKKNIQOOSYSNINNLEFEGYFFKVMKLDKDEAKWKLMEINIRKKNEFSN 300
DB 633 WNSIKKLFTRKKNNVPOPYTININNLEFEGYFFKVMKLDKDEAKWKLMEINIRKKNEFSN 692
QY 301 LENNRDYLENAIELLDHLKETATTCNDNTNACETSHNATNP 345
DB 693 LENNRDYLENAIELLDHLKETATTCNDNTNACETSHNATNP 737

RESULT 7
Q8T6K4 PRELIMINARY; PRT; 465 AA.
AC Q8T6K4;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-M57;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT "Identification of a conserved Plasmodium falciparum var gene
 RT implicated in malaria in pregnancy."
 RL J. Infect. Dis. 0:0-0(2002).
 DR EMBL; AF458892; AAL84275.1; -
 FT NON_TER 1 1
 FT 465 465
 SQ SEQUENCE 465 AA; 55496 MW; 6B0E35F62468A37B CRC64;

Query Match 85.0%; Score 1631; DB 5; Length 465;
 Best Local Similarity 98.0%; Pred. No. 2.3e-91;
 Matches 290; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 YKEIQSYLSNDNKEFVNINSEYKQFYKELKETQYATNDPFLNLINSGKYCKGGLPGEK 60
 |||||:|||||
 DB 170 YKEIESYVSNDAKRVNNINSEYKQFYKELKETQYATNDPFLNLINSGKYCKGGLPGEK 229
 |||||:|||||
 OY 61 DITFTNSADDKGIFRSEYCYQVPCGVCAGDGIKTHKSDNDRERNNEDYKPPMGVAKPT 120
 |||||:|||||
 DB 230 DITFTNSADDKGIFRSEYCYQVPCGVCAGDGIKTHKSDNDRERNNEDYKPPMGVAKPT 289
 |||||:|||||
 OY 121 NITVLYSGNEOGDITOKLENFCSSTNYKDKNNQKWECCYKDEINRCKLEQNTETINNND 180
 |||||:|||||
 DB 290 NITVLYSGNEOGDITOKLENFCSSTNYKDKNNQKWECCYKDEINRCKLEQNTETINNND 349
 |||||:|||||
 OY 181 PRTISFHNFEFLWVYLLRDITKMDKLTCTCINNTHHCIDECNRNCLCFDRWVQKKEE 240
 |||||:|||||
 DB 350 PRTISFHNFEFLWVYLLRDITKMDKLTCTCINNTHHCIDECNRNCLCFDRWVQKKEE 409
 |||||:|||||
 OY 241 WNSIKKLFTRKKNNIOOSYYSNINLFEgyFFKVMOKLDEAKKWEKMEINIRKK 296
 |||||:|||||
 DB 410 WNSIKKLFTRKKNNIOOSYYSNINLFEgyFFKVMOKLDEAKKWEKMEINIRKK 465
 |||||:|||||

RESULT 8
 OBT6K3 PRELIMINARY; PRT; 461 AA.
 AC OBT6K3;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Erythrocyte membrane protein 1 (Fragment).
 VAR.
 Plasmodium falciparum.
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-M58;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT "Identification of a conserved Plasmodium falciparum var gene
 RT implicated in malaria in pregnancy."
 RL J. Infect. Dis. 0:0-0(2002).
 DR EMBL; AF458893; AAL84276.1; -
 FT NON_TER 1 1
 FT 461 461
 SQ SEQUENCE 461 AA; 54946 MW; F4C8863D50DA50EF CRC64;

Query Match 84.7%; Score 1625; DB 5; Length 461;
 Best Local Similarity 98.0%; Pred. No. 5.2e-91;
 Matches 289; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 YKEIQSYLSNDNKEFVNINSEYKQFYKELKETQYATNDPFLNLINSGKYCKGGLPGEK 60
 |||||:|||||
 DB 167 YKEIESYVSNDAKRVNNINSEYKQFYKELKETQYATNDPFLNLINSGKYCKGGLPGEK 226
 |||||:|||||
 OY 61 DITFTNSADDKGIFRSEYCYQVPCGVCAGDGIKTHKSDNDRERNNEDYKPPMGVAKPT 120
 |||||:|||||

DB 227 DITFTNSADDKGIFRSEYCYQVPCGVCAGDGIKTHKSDNDRERNNEDYKPPMGVAKPT 286
 |||||:|||||
 OY 121 NITVLYSGNEOGDITOKLENFCSSTNYKDKNNQKWECCYKDEINRCKLEQNTETINNND 180
 |||||:|||||
 DB 287 NITVLYSGNEOGDITOKLENFCSSTNYKDKNNQKWECCYKDEINRCKLEQNTETINNND 346
 |||||:|||||
 OY 181 PRTISFHNFEFLWVYLLRDITKMDKLTCTCINNTHHCIDECNRNCLCFDRWVQKKEE 240
 |||||:|||||
 DB 347 PRTISFHNFEFLWVYLLRDITKMDKLTCTCINNTHHCIDECNRNCLCFDRWVQKKEE 406
 |||||:|||||
 OY 241 WNSIKKLFTRKKNNIOOSYYSNINLFEgyFFKVMOKLDEAKKWEKMEINIRKK 295
 |||||:|||||
 DB 407 WNSIKKLFTRKKNNIOOSYYSNINLFEgyFFKVMOKLDEAKKWEKMEINIRKK 461
 |||||:|||||

Query Match 83.9%; Score 1610; DB 5; Length 492;
 Best Local Similarity 89.2%; Pred. No. 4.5e-90;
 Matches 288; Conservative 17; Mismatches 18; Indels 0; Gaps 0;

OY 1 YKEIQSYLSNDNKEFVNINSEYKQFYKELKETQYATNDPFLNLINSGKYCKGGLPGEK 60
 |||||:|||||
 DB 170 YKEIESYVSNDAKRVNNINSEYKQFYKELKETQYATNDPFLNLINSGKYCKGGLPGEK 229
 |||||:|||||
 OY 61 DITFTNSADDKGIFRSEYCYQVPCGVCAGDGIKTHKSDNDRERNNEDYKPPMGVAKPT 120
 |||||:|||||
 DB 230 DITFTNSADDKGIFRSEYCYQVPCGVCAGDGIKTHKSDNDRERNNEDYKPPMGVAKPT 289
 |||||:|||||
 OY 121 NITVLYSGNEOGDITOKLENFCSSTNYKDKNNQKWECCYKDEINRCKLEQNTETINNND 180
 |||||:|||||
 DB 290 NITVLYSGNEOGDITOKLENFCSSTNYKDKNNQKWECCYKDEINRCKLEQNTETINNND 349
 |||||:|||||
 OY 181 PRTISFHNFEFLWVYLLRDITKMDKLTCTCINNTHHCIDECNRNCLCFDRWVQKKEE 240
 |||||:|||||
 DB 350 PRTISFHNFEFLWVYLLRDITKMDKLTCTCINNTHHCIDECNRNCLCFDRWVQKKEE 409
 |||||:|||||
 OY 241 WNSIKKLFTRKKNNIOOSYYSNINLFEgyFFKVMOKLDEAKKWEKMEINIRKKNEFSN 300
 |||||:|||||
 DB 410 WNSIKKLFTRKKNNIOOSYYSNINLFEgyFFKVMOKLDEAKKWEKMEINIRKKNEFSN 469
 |||||:|||||
 OY 301 LENNDYLENAIELLDHLKETA 323
 |||||:|||||
 DB 470 LENNDYLENAIELLDHLKETA 492
 |||||:|||||

RESULT 10
 OBT6K0 PRELIMINARY; PRT; 494 AA.
 AC OBT6K0;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (Tremblrel. 21, last sequence update)
 DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
 GN Erythrocyte membrane protein 1 (Fragment).
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VAN434;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT "Identification of a conserved Plasmodium falciparum var gene
 RT implicated in malaria in pregnancy.";
 RL J. Infect. Dis. 0:0-0(2002).
 DR EMBL: AF458896; AAL84279.1; -.
 FT NON_TER 1 1
 FT NON_TER 494 494
 SQ SEQUENCE 494 AA; 59240 MW; 739BE121652A367B CRC64;

Query Match 83.9%; Score 1609; DB 5; Length 494;
 Best Local Similarity 88.9%; Pred. No. 5.2e-90;
 Matches 288; Conservative 14; Mismatches 22; Indels 0; Gaps 0;

OY 1 YEKEIOSYLSNDKFNANNINSEYKQFEKLEQVATNDPFLNLNEGKCKGGLPGEK 60
 DB 171 YEKEIOSYLSNDKFNANNINSEYKQFEKLEQVATNDPFLNLNEGKCKGGLPGEK 230
 OY 61 DIFTNSADKGIFFRSEYQVCPDGVKCDGKITYTHKSDNDRVNNEDYKPPWVKPT 120
 DB 231 DIFTNSADKGIFFRSEYQVCPDGVKCDGKITYTHKSDNDRVNNEDYKPPWVKPT 290
 OY 121 NITVLYSGNQGDIOTKLENFCSNSTYKDKNNQKWCYKDEINCKLEQNTIENNDN 180
 DB 291 NITVLYSGNQGDIOTKLENFCSNSTYKDKNNQKWCYKDEINCKLEQNTIENNDN 350
 OY 181 PKIISFNFPFLWVYLLRDTIKNDKLCINNTTHCHIDECKRNLCEDRWYKQKEE 240
 DB 351 PKIISFNFPFLWVYLLRDTIKNDKLCINNTTHCHIDECKRNLCEDRWYKQKEE 410
 OY 241 WNSIKKLFYTKKNNIQOYSYNNINLFEYGFYKVDKLDKDEAKKELMENIKRKNFEFSN 300
 DB 411 WNSIKKLFYTKKNNIQOYSYNNINLFEYGFYKVDKLDKDEAKKELMENIKRKNFEFSN 470
 OY 301 LENNRDYLENAIELLDLHLEKETAT 324
 DB 471 LENNRDYLENAIELLDLHLEKETAT 494

Query Match 83.9%; Score 1609; DB 5; Length 494;
 Best Local Similarity 88.9%; Pred. No. 5.2e-90;
 Matches 288; Conservative 14; Mismatches 22; Indels 0; Gaps 0;

DT 01-JUN-2002 (Tremblrel. 21, last sequence update)
 DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
 GN Erythrocyte membrane protein 1 (Fragment).
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VAN434;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT "Identification of a conserved Plasmodium falciparum var gene
 RT implicated in malaria in pregnancy.";
 RL J. Infect. Dis. 0:0-0(2002).
 DR EMBL: AF458896; AAL84279.1; -.
 FT NON_TER 1 1
 FT NON_TER 494 494
 SQ SEQUENCE 494 AA; 59240 MW; 739BE121652A367B CRC64;

Query Match 83.9%; Score 1609; DB 5; Length 494;
 Best Local Similarity 87.7%; Pred. No. 1.2e-89;

Matches 286; Conservative 16; Mismatches 24; Indels 0; Gaps 0;

OY 1 YEKEIOSYLSNDKFNANNINSEYKQFEKLEQVATNDPFLNLNEGKCKGGLPGEK 60
 DB 171 YEKEIOSYLSNDKFNANNINSEYKQFEKLEQVATNDPFLNLNEGKCKGGLPGEK 230
 OY 61 DIFTNSADKGIFFRSEYQVCPDGVKCDGKITYTHKSDNDRVNNEDYKPPWVKPT 120
 DB 231 DIFTNSADKGIFFRSEYQVCPDGVKCDGKITYTHKSDNDRVNNEDYKPPWVKPT 290
 OY 121 NITVLYSGNQGDIOTKLENFCSNSTYKDKNNQKWCYKDEINCKLEQNTIENNDN 180
 DB 291 NITVLYSGNQGDIOTKLENFCSNSTYKDKNNQKWCYKDEINCKLEQNTIENNDN 350
 OY 181 PKIISFNFPFLWVYLLRDTIKNDKLCINNTTHCHIDECKRNLCEDRWYKQKEE 240
 DB 351 PKIISFNFPFLWVYLLRDTIKNDKLCINNTTHCHIDECKRNLCEDRWYKQKEE 410
 OY 241 WNSIKKLFYTKKNNIQOYSYNNINLFEYGFYKVDKLDKDEAKKELMENIKRKNFEFSN 300
 DB 411 WNSIKKLFYTKKNNIQOYSYNNINLFEYGFYKVDKLDKDEAKKELMENIKRKNFEFSN 470
 OY 301 LENNRDYLENAIELLDLHLEKETAT 326
 DB 471 LENNRDYLENAIELLDLHLEKETAT 496

Query Match 82.2%; Score 1577; DB 5; Length 494;
 Best Local Similarity 87.6%; Pred. No. 4.5e-88;
 Matches 283; Conservative 14; Mismatches 26; Indels 0; Gaps 0;

DT 01-JUN-2002 (Tremblrel. 21, last sequence update)
 DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
 GN Erythrocyte membrane protein 1 (Fragment).
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K117;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT "Identification of a conserved Plasmodium falciparum var gene
 RT implicated in malaria in pregnancy.";
 RL J. Infect. Dis. 0:0-0(2002).
 DR EMBL: AF458891; AAL84274.1; -.
 FT NON_TER 1 1
 FT NON_TER 494 494
 SQ SEQUENCE 494 AA; 58870 MW; 56A34264A9285321 CRC64;

Query Match 82.2%; Score 1577; DB 5; Length 494;
 Best Local Similarity 87.6%; Pred. No. 4.5e-88;
 Matches 283; Conservative 14; Mismatches 26; Indels 0; Gaps 0;

OY 1 YEKEIOSYLSNDKFNANNINSEYKQFEKLEQVATNDPFLNLNEGKCKGGLPGEK 60
 DB 172 YEKEIOSYLSNDKFNANNINSEYKQFEKLEQVATNDPFLNLNEGKCKGGLPGEK 231
 OY 61 DIFTNSADKGIFFRSEYQVCPDGVKCDGKITYTHKSDNDRVNNEDYKPPWVKPT 120
 DB 232 DIFTNSADKGIFFRSEYQVCPDGVKCDGKITYTHKSDNDRVNNEDYKPPWVKPT 291
 OY 121 NITVLYSGNQGDIOTKLENFCSNSTYKDKNNQKWCYKDEINCKLEQNTIENNDN 180
 DB 292 NITVLYSGNQGDIOTKLENFCSNSTYKDKNNQKWCYKDEINCKLEQNTIENNDN 351
 OY 181 PKIISFNFPFLWVYLLRDTIKNDKLCINNTTHCHIDECKRNLCEDRWYKQKEE 240
 DB 352 PKIISFNFPFLWVYLLRDTIKNDKLCINNTTHCHIDECKRNLCEDRWYKQKEE 411
 OY 241 WNSIKKLFYTKKNNIQOYSYNNINLFEYGFYKVDKLDKDEAKKELMENIKRKNFEFSN 300
 DB 411 WNSIKKLFYTKKNNIQOYSYNNINLFEYGFYKVDKLDKDEAKKELMENIKRKNFEFSN 470

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 3.48066 Seconds

(without alignments)
2916.372 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747

Perfect score: 1918

Sequence: 1-YEKRIQSYLSNDKRFVNNIN.....CRDNTNACETSHNATNP 345

Scoring table: BLOSUM62

Searched: 262574 segs, 29422922 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Issued_Patents_AA.*

- 1: /cg2_6/ptodata/1/1aa/5A_COMB.pep.*
- 2: /cg2_6/ptodata/1/1aa/5B_COMB.pep.*
- 3: /cg2_6/ptodata/1/1aa/5A_COMB.pep.*
- 4: /cg2_6/ptodata/1/1aa/5B_COMB.pep.*
- 5: /cg2_6/ptodata/1/1aa/5A_COMB.pep.*
- 6: /cg2_6/ptodata/1/1aa/5B_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	394.5	20.6	2710	2	US-08-568-459A-12
2	394.5	20.6	2710	2	US-08-487-826B-12
3	394.5	20.6	2710	4	US-09-210-288-12
4	394.5	20.6	3060	2	US-08-487-826B-14
5	380	19.8	700	2	US-08-568-459A-10
6	380	19.8	700	2	US-08-487-826B-10
7	380	19.8	700	4	US-09-210-288-10
8	380	19.8	2182	2	US-08-487-826B-16
9	231	12.0	921	2	US-08-568-459A-8
10	231	12.0	921	2	US-08-487-826B-8
11	231	12.0	921	4	US-09-210-288-8
12	145	7.6	749	2	US-08-568-459A-6
13	145	7.6	749	2	US-08-487-826B-6
14	145	7.6	749	4	US-09-210-288-6
15	137	7.1	1435	2	US-08-568-459A-4
16	137	7.1	1435	2	US-08-487-826B-4
17	137	7.1	1435	4	US-09-210-288-4
18	136	7.1	1147	3	US-08-470-260-5
19	136	7.1	1147	3	US-08-471-491-5
20	136	7.1	1147	4	US-08-466-662-5
21	136	7.1	3289	2	US-08-477-451-2
22	123	6.4	2391	2	US-08-446-855A-2
23	123	6.4	2391	4	US-09-150-741-2
24	120	6.3	463	4	US-08-845-258-25
25	120	6.3	463	4	US-08-990-571-25
26	120	6.3	463	4	US-08-723-142A-25
27	120	6.3	463	4	US-09-528-784A-25

28	118.5	6.2	1085	1	US-08-431-080-28	Sequence 28, Appl
29	118.5	6.2	1085	2	US-08-938-534-28	Sequence 28, Appl
30	118.5	6.2	1085	4	US-09-345-294-28	Sequence 28, Appl
31	118	6.2	1115	2	US-08-568-459A-2	Sequence 2, Appl1
32	118	6.2	1115	2	US-08-487-826B-2	Sequence 2, Appl1
33	118	6.2	1115	4	US-09-210-288-2	Sequence 2, Appl1
34	118	6.2	1115	6	5198347-6	Patent No. 5198347
35	116.5	6.1	290	4	US-08-903-801-1	Sequence 1, Appl1
36	116.5	6.1	290	4	US-09-295-055-1	Sequence 1, Appl1
37	114.5	6.0	1104	4	US-08-923-992A-4	Sequence 4, Appl1
38	112.5	5.9	1151	4	US-09-134-001C-3242	Sequence 3242, Ap
39	112	5.8	10182	4	US-08-923-992A-2	Sequence 3159, Ap
40	110.5	5.8	1164	4	US-08-923-992A-2	Sequence 10, Appl
41	110.5	5.8	1164	4	US-08-923-992A-10	Sequence 2, Appl1
42	110	5.7	778	6	5198347-4	Patent No. 5198347
43	108	5.6	3135	1	US-08-323-170B-2	Sequence 2, Appl1
44	108	5.6	3135	4	US-08-954-441-2	Sequence 2, Appl1
45	107.5	5.6	599	2	US-08-910-551B-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-08-568-459A-12
Sequence 12, Application US/08568459A
Patent No. 5849306

GENERAL INFORMATION:

APPLICANT: Sim, Kim I.
APPLICANT: Chituls, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:

ORGANISM: Plasmodium falciparum
US-08-568-459A-12

Query Match 20.6%; Score 394.5; DB 2; Length 2710;

TELEFAX: (619) 235-0170

MEDIUM TYPE: Floppy disk

COMPUTER READABLE;

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/210,288
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Fuller, Michael
 REGISTRATION NUMBER: 36,516
 REFERENCE/DOCKET NUMBER: NIH21.1FWDY1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2710 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium falciparum
 US-09-210-288-12

US-08-568-459A-10
: Sequence 10, Application US/08568459A
: Patent No. 5849306
: GENERAL INFORMATION:
: APPLICANT: Sim, Kim L.
: APPLICANT: Chitnis, Chetan
: APPLICANT: Miller, Louis H.
: APPLICANT: Peterson, David S.
: APPLICANT: Su, Xin-zhaun
: APPLICANT: Wellens, Thomas E.
: TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
: TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knodbe Martens Olson & Bear
: STREET: 620 Newport Center Drive 16th Floor
: CITY: Newport Beach
: STATE: California
: COUNTRY: US
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/568,459A
: FILING DATE: 07-DEC-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelien, Ned
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER: NIH21.001CPI
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 700 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Plasmodium falciparum
US-08-568-459A-10

Query Match 19.8%; Score 380; DB 2; Length 700;
Best Local Similarity 28.7%; Pred. No. 1.3e-24;
Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

1 YEREI-----QSYLSNDKFFVNNINSEYKOFYKLETOYAINDTFL 43
285 YETELISGGGSGSPKRTKRAARSSSSDDN-----GYESKFYKLEKVEGYQVDVDFL 336
44 NLNKGKCYCG---GLPGEKDIPTNSADDKGIFRSEYCOVCPGCGYKCGIKYTHKS 99
337 KLINKEGICQKOPQYGNKADNDVDFNTEKYK-TESRTEICEPCWCGLEKGGPPMKYVG 395
100 DNDREYNNEDYKPPGVKFTNITVLYSGNEGDIYOKLEFNCNSSTNYKKNOKWECY 159
396 DKTGSAKTYDPK---NITDIPVLYPDKSQOONILKRYKNCEKCAP-GGQIKKMGY 451
160 YKD-----ENINRCKLEONTEINNDNPKLISFHNFEFLAVYLLMDITKMDKLTCTINN 214
452 YDEHPPSSKNNNNVEGATWDFKTOGKQYKYNFPMVDMDLHSVETELSKCINN 511
215 TTT-----HCIDECNRNCICFDRWYKQKEEENSJKLFTKKNNIQOYSTYNNINLFESEY 270
512 NTNGATCRNNKCKTKDCCCFKRWVEKKQOEWMAIKDFGKQDTIVQO-----KGLIYFSP 566
271 FKVMKLDKDEAKWELMENIKRKNESNLENNRDYLE-----NAIELL 316

US-08-487-826B-10
: Sequence 10, Application US/08487826B
: Patent No. 5993827
: GENERAL INFORMATION:
: APPLICANT: Sim, Kim L.
: APPLICANT: Chitnis, Chetan
: APPLICANT: Miller, Louis H.
: APPLICANT: Peterson, David S.
: APPLICANT: Su, Xin-zhaun
: APPLICANT: Wellens, Thomas E.
: TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
: TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knodbe Martens Olson & Bear
: STREET: 620 Newport Center Drive 16th Floor
: CITY: Newport Beach
: STATE: California
: COUNTRY: US
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/487,826B
: FILING DATE: 10-SEP-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelien, Ned
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER: NIH21.001CPI
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 700 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Plasmodium falciparum
US-08-487-826B-10

Query Match 19.8%; Score 380; DB 2; Length 700;
Best Local Similarity 28.7%; Pred. No. 1.3e-24;
Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

1 YEREI-----QSYLSNDKFFVNNINSEYKOFYKLETOYAINDTFL 43
285 YETELISGGGSGSPKRTKRAARSSSSDDN-----GYESKFYKLEKVEGYQVDVDFL 336
44 NLNKGKCYCG---GLPGEKDIPTNSADDKGIFRSEYCOVCPGCGYKCGIKYTHKS 99
337 KLINKEGICQKOPQYGNKADNDVDFNTEKYK-TESRTEICEPCWCGLEKGGPPMKYVG 395
100 DNDREYNNEDYKPPGVKFTNITVLYSGNEGDIYOKLEFNCNSSTNYKKNOKWECY 159
396 DKTGSAKTYDPK---NITDIPVLYPDKSQOONILKRYKNCEKCAP-GGQIKKMGY 451

```

QY 160 YKD-----ENINCKLEQNTNNDNPKIISFHFELWYTLIRDTIKNDKLTICIN 214
D 452 YDEHRSKNNNNVCBETWMTKFTOGKOTVKSYNFFWDMVHDLHDSVEKTELKSCIN 511
QY 215 TTT-----HCIDECNRNLCFDRWVKQKEEEMNSIKLFTKKNIOOSYSINNNLFEQYF 270
D 512 NTNGNCRNNKCKTDCGCKQKWEKKQEMAIKDHFGKOTDVOO-----KGLIYFSP 566
QY 271 FKVDKLDKDEARKWELMENIKRKNNEFSNLNRRDYLE-----NAIFELL 316
D 567 YGVLDLV-----LKGNNLQNIKDVHGDTDDIKHKKLLDEDAVAVVVGKNDTTIDKLL 622
QY 317 DHIKETATICKDNNTNEACE 336
D 623 QHEKEQAEOCKOK--OECE 640

RESULT 7
US-09-210-288-10
Sequence 10, Application US/09210288
Patent No. 6392026
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210, 288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 700 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-10

Query Match 19.8%; Score 380; DB 4; Length 700;
Best Local Similarity 28.7%; Pred. No. 1.3e-24;
Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;
QY 1 YEKEI-----OSYLSNDKRFVNNINSEYKQYERKLEKTOYATNDPFL 43
D 285 YETELSGGSGKSPKRTKRAARSSSSSDN-----GYESKRYKKLKEGYGVYDVVKFL 336

```

```

QY 44 NLNEGKYCKG---GLPGEKDTITNSADDKGIYRSEYQVCPDCKGKIGIKYTHKS 99
D 337 KILNKEGICQKQPOVNGEKNADNDVETNEKYYK--TFSRTICEPCPMGLKRGKGPWKVG 395
QY 100 DNDREKRNNDYKPPKAVKRTNITVLSGNDGGDITQKLENFCSNINADKNQKKECY 159
D 396 DKTCGSAKITYDPK---NITDIPVLVDPKSOQNLKRYKRFCEKGP--GGQIKKWQCY 451
QY 160 YKD-----ENINCKLEQNTNNDNPKIISFHFELWYTLIRDTIKNDKLTICIN 214
D 452 YDEHRSKNNNNVCBETWMTKFTOGKOTVKSYNFFWDMVHDLHDSVEKTELKSCIN 511
QY 215 TTT-----HCIDECNRNLCFDRWVKQKEEEMNSIKLFTKKNIOOSYSINNNLFEQYF 270
D 512 NTNGNCRNNKCKTDCGCKQKWEKKQEMAIKDHFGKOTDVOO-----KGLIYFSP 566
QY 271 FKVDKLDKDEARKWELMENIKRKNNEFSNLNRRDYLE-----NAIFELL 316
D 567 YGVLDLV-----LKGNNLQNIKDVHGDTDDIKHKKLLDEDAVAVVVGKNDTTIDKLL 622
QY 317 DHIKETATICKDNNTNEACE 336
D 623 QHEKEQAEOCKOK--OECE 640

RESULT 8
US-08-487-826B-16
Sequence 16, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487, 826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2182 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal

```

US-08-487-826B-16

Query Match 19.8%; Score 380; DB 2; Length 2182;
Best Local Similarity 28.7%; Pred. No. 5,9e-24;
Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

QY 1 YEKEI-----QSYLSNDKKEFVNNINSEYKQFYKELKETOYATNDTEL 43
DB 413 YETETISGGGSGKSPKRTKRAARSSSSDDN-----GYESKFKYKLEKVGQYDVDFKL 464
QY 44 NLNEGKCYGCK-----GLPGEKDTFTNSADBDKGFYRSEYCGYPCGVCDCIKTHKS 99
DB 465 KILANKGICQKQPOVGNENKADNDFTNEKYK-TFSRTLECEPCWCGLEKGGPPKVVG 523
QY 100 DNDREVRNEDYKPPWGVKFTNITVLVSGNEQDITQKLEFNCSTNYKDKKNQKWCY 159
DB 524 DKTCGSAKTKTYDPK---NTDIPVLVPRKSGQNLKTKYKNCEKAP-GGGQIKRMQVC 579
QY 160 YKD-----ENINCKLEQNTNINNDPKIISFNFELWVYTLIRPTIKNDKLTQICNN 214
DB 580 YEHRRSSKNNNVCVEGTDKFTQGRQYKSYNVFPMVDVHDLHDSVEKTELKSCINN 639
QY 215 TTT---HCIDECNRNCLCFDRVWVKOEEMNSIKKLTFRKKNIQOOSYNNINLEGEY 270
DB 640 NNGNNTCRNNKCKTDCGCFQKWEKKQDPMALIKHFGKQDIDVQO-----KGLVFSB 694
QY 271 FVYMDKLDEAKWELMENIKRKNFSLNENRDYLE-----NALELL 316
DB 695 YGVLIDY---LKGMLLONIKVDHGDITDIKIKILDEDAVAVVLGSKDWTIDKLL 750
QY 317 DHKETFATICKDNNTWEACE 336
DB 751 QHEKEQAEQCKOK--OECE 768

RESULT 9

US-08-568-459A-8
Sequence 8, Application US/08568459A
Patent No. 5849306

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Israelien, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:

ORGANISM: Plasmodium falciparum
US-08-568-459A-8

Query Match 12.0%; Score 231; DB 2; Length 921;
Best Local Similarity 24.1%; Pred. No. 1.3e-11;
Matches 91; Conservative 55; Mismatches 147; Indels 84; Gaps 17;

QY 36 YATNDFLLNLDEGKYCKGLGKEKDTFTNSADBDKGFYRSEYQVCP---DCGVKCD 91
DB 6 YTSIGLEFNSLSKSGPCKKNDNAEDNIDF---GDECKTFKREADNCKPQSQFTVDC-KKCN 61
QY 92 GIKYTHKSDNDRVRNEDYKPPWGVKFTNITVLVSGNEQDITQKLE-----NF 141
DB 62 GDDTKGCKNGSKKNGDY-----ITASDIENGNSIGNIDVYSDKANGFNGLDA 114
QY 142 CNSTNYKDKKNQKWCYKDEININCKLEQ-NTEINNDNPKTISFNFELWVYTLIR 200
DB 115 CGSANIFKGRKEQWKC-AKVGGLDVCGLKNGSIDKQKQIIIRALKRWVEYFLD 173
QY 201 TIKMDKLKTCI--NNTTHCIDECNRNCLCFDRVWVKOEEMNSIKKLF-TKKNNIOOS 257
DB 174 YKINAKISHCYKKNENST-CYNDPCNKCTVEEMINOKRTEWNIKIKHYKTONENGNN 232
QY 258 YYSNNINL-----FEGYFKVWDKLDKDEAKWKE-----LM 288
DB 233 KSLVTDILGALPQSDVNAKIPCSGLTAFPS--FCGINGADNSEKKEGEDYDLVLM 290
QY 288 ENIK-----KKNEFSNLEN-----NDYLENAIEILLDLKETATLC-----KD 328
DB 291 KMLEQIOECKKKHGETSVENGKSCPTLDNTTLEETPE--EENQVAPNIPCQVTEW 348
QY 329 NNTNEACETSHATNP 345
DB 349 KKKEEETCTPASPVP 365

RESULT 10

US-08-487-826B-8
Sequence 8, Application US/08487826B
Patent No. 5893827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-487-826B-8

Query Match 12.0%; Score 231; DB 2; Length 921;
Best Local Similarity 24.1%; Pred. No. 1,3e-11;
Matches 91; Conservative 55; Mismatches 147; Indels 84; Gaps 17;

QY 36 YATNDTFLNLNBSKYGKGLPGEKDTFTNSADKGIIFYRSEYCOYCP----DCGYKCD 91
DB 6 YTSIGLFLNLSKSGPCCKDNDNAEDNIDF---GDEGTFKFEADCKCSQFTYDC-KNCN 61
QY 92 GIKYTHKSDNDREVRNEDYKPPWGVKPTNITVLYSGNEGDTQKLE-----NF 141
DB 62 GGDYKKGKNGSKNGKKNNDY-----ITASDIENGNSIGNIDMVSDDKANGFNGLDA 114
QY 142 CNSSTYKDKNNQWECYVDENINRCKLEQ-NTEINNDNPKIISFNFEFLWYTYLLRD 200
DB 115 CGSANITKGIKREQWKC-AKYCGLDVGLKNGSIDKQKQIITIRALLKRWVEYLED 173
QY 201 TIKWMDLKTICI--NNTTTCIDECNRCICFDWVVKQKEEWSIKKLF-TKKNNIOOS 257
DB 174 YKNTNAKISHCTKKNDEST-CTNDCPNKCTCVEEMINQKTEWNIKKHYKTQNGNDN 232
QY 258 YYSNNNL-----FEGYFFVMDKLDKDEAKWKE-----LM 288
DB 233 MSLVTDILGALQPSDVNKAIRKPCSLTAPES--FCGLNGADNSEKKEGEDYDVLVCLM 290
QY 289 ENIK-----RKNEFSNLEN-----NRDYLENAIELLDHLKETATIC-----KD 328
DB 291 KMLEKQIOECKKKHGGTSVENGSKSTPLDNTLTLEEPPIE--EENOVEAPNICQKQVED 348
QY 329 NNTNEACETSHNATTNP 345
DB 349 KKKEEBETCTPASVP 365

RESULT 11
US-09-210-288-8
Sequence 8, Application US/09210288
Patent No. 6392026
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US

ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-8

Query Match 12.0%; Score 231; DB 4; Length 921;
Best Local Similarity 24.1%; Pred. No. 1,3e-11;
Matches 91; Conservative 55; Mismatches 147; Indels 84; Gaps 17;

QY 36 YATNDTFLNLNBSKYGKGLPGEKDTFTNSADKGIIFYRSEYCOYCP----DCGYKCD 91
DB 6 YTSIGLFLNLSKSGPCCKDNDNAEDNIDF---GDEGTFKFEADCKCSQFTYDC-KNCN 61
QY 92 GIKYTHKSDNDREVRNEDYKPPWGVKPTNITVLYSGNEGDTQKLE-----NF 141
DB 62 GGDYKKGKNGSKNGKKNNDY-----ITASDIENGNSIGNIDMVSDDKANGFNGLDA 114
QY 142 CNSSTYKDKNNQWECYVDENINRCKLEQ-NTEINNDNPKIISFNFEFLWYTYLLRD 200
DB 115 CGSANITKGIKREQWKC-AKYCGLDVGLKNGSIDKQKQIITIRALLKRWVEYLED 173
QY 201 TIKWMDLKTICI--NNTTTCIDECNRCICFDWVVKQKEEWSIKKLF-TKKNNIOOS 257
DB 174 YKNTNAKISHCTKKNDEST-CTNDCPNKCTCVEEMINQKTEWNIKKHYKTQNGNDN 232
QY 258 YYSNNNL-----FEGYFFVMDKLDKDEAKWKE-----LM 288
DB 233 MSLVTDILGALQPSDVNKAIRKPCSLTAPES--FCGLNGADNSEKKEGEDYDVLVCLM 290
QY 289 ENIK-----RKNEFSNLEN-----NRDYLENAIELLDHLKETATIC-----KD 328
DB 291 KMLEKQIOECKKKHGGTSVENGSKSTPLDNTLTLEEPPIE--EENOVEAPNICQKQVED 348
QY 329 NNTNEACETSHNATTNP 345
DB 349 KKKEEBETCTPASVP 365

RESULT 12
US-08-568-459A-6
Sequence 6, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.

APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-6

Query Match 7.6%; Score 145; DB 4; Length 749;
Best Local Similarity 21.5%; Pred. No. 0.00026;
Matches 61; Conservative 46; Mismatches 103; Indels 74; Gaps 11;

QY 91 DGIR-YTHKSDNDRERNNDYKRPVGVKPTNITVLYSGNEOGDITQKLEFNCSSYNYK 149
D 162 EGLADHINKANAYEAMHLKREYENAGDGKICN-AIIGSYADIGDIYVGLDWMRDINTN-- 218
D 150 DKNNAKECYKIDENIRCLLEONTETIN-----DNPKTISHNFPEL 192
D 219 -KISEKQKLFMGGSNRKQNDNNERNKWKORNLWSSMVKHPRKGTCKRHNFEEK 277
QY 193 WVTYLLDITIKMNDKLT-----KTCINNTTHICDECRNCLCFDRWVKOKEE 240
D 278 -IQPLFKWIKEMGDPEEKEGTEKQLEKICEKNKCE--KCCNACSSTEKMKIKERKE 334
QY 241 WNSIKLFTKKKNIQOQSYYSINNNLEGEYFEKVDKLDKDEAKKELMENIKRKNNEFSN 300
D 335 YNLSQSKFSDSKLNRK-----NNLYN-----KREDSKAVLRSESKQCSN 374
QY 301 LENNROYLENAIELLDHLKETATICDNNNTNEACETSHNATN 344
D 375 IEFN-----DETFTF--PNKYKCAVCNPPSSS 401

RESULT 15
US-08-568-459A-4
Sequence 4, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnais, Chetan
APPLICANT: Miller, Louis H.

APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-4

Query Match 7.1%; Score 137; DB 2; Length 1435;
Best Local Similarity 20.5%; Pred. No. 0.003;
Matches 86; Conservative 57; Mismatches 149; Indels 128; Gaps 19;

QY 1 YEKIQSYLSNDN-----KFVN-NINSEYKOFEXKRETOYATVNDLFNLNGKXCK- 53
D 75 FNRKRSYIGPDNIDKMSLINKNNEMNNYQSLSTS-----SLIKONKIVPI 126
QY 54 GGLPERDITFTNSADKGIYRSEYQVCPDGVKCDGIKYTHKSDNDRER-VNEDYK 112
D 127 NAVRVSILSFLDSRINNG-RNTSSNNEVLSNCKEKKRGMKWDCKKKNDRSNYICIDRR 185
QY 113 PPGVGVKPTNITVLYSGNEOGD-----ITQKLEN-----FQNSSTN----- 147
D 186 IQLCIVNLSTIKTYKTMKDIHLEASKRESQLLKKNNDNNYNSKPCNDLKNSLFYGH 245
QY 148 -----YKDKNNOK-----WECYKRD--E 163
D 246 AMGNDMPGCGYSTAEKNKIQVFNGANGSEIKKFNKRWMDPREFKLEAMLSERKN 305
QY 164 NINRCKLEONTENNDNPKIISFNHFEWVYLLRDTIKWMDLKTICINNTTTH-CIDE 222
D 306 NINCKNIPQEL-QITQWIKEMHGEF-----LLERDNRSLPKSK-CKNNTLYEACEKE 358
QY 223 CNRNCLOFDRWVKOKEEMNSIKLFTKKKNIQOQSYYSINNNLEGEYFEKVMKLDDEA 282
D 359 CIDPCMKYRDMITRSKEWHLISKEYETQKVPKEN-----AENYLKISR--NKND 408
QY 283 KWEKLENIKRRKNNEFSNLENNRDLLENALIELLDHLKETATICK-----DNNNTNEAC 336
D 409 KVSLLANCAEYKYCD-----CKHTTTLVKSVLNGNDNTIKERRE 450

Mon Apr 28 13:47:52 2003

us-10-087-013-2_copy_403_747.rai

Page 10

Search completed: April 28, 2003, 10:32:06
Job time : 9.48066 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 . Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:32:15 ; Search time 7.70718 Seconds
(without alignments)
3586.886 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747

Perfect score: 1918

Sequence: 1 YKEKTSYLSNDKFNIN.....CKDNTNACETSHNATNP 345

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

tal number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	394.5	20.6	2710	9	US-10-153-273-12
2	380	19.8	700	9	US-10-153-273-10
3	231	12.0	921	9	US-10-153-273-8
4	162	8.4	807	9	US-09-820-843A-108
5	161	8.4	972	10	US-09-924-154-16
6	151.5	7.9	1086	10	US-09-924-154-15
7	147	7.7	1421	10	US-09-924-154-13
8	145	7.6	749	9	US-10-153-273-6
9	144	7.5	861	9	US-09-820-843A-109
10	137	7.1	1435	9	US-10-153-273-4
11	136	7.1	1338	10	US-09-402-100-4
12	133	6.7	1501	10	US-09-924-154-17
13	128	6.7	1143	10	US-09-924-154-14
14	122	6.4	1805	9	US-09-820-843A-73
15	120	6.3	463	9	US-09-286-488-25
16	120	6.3	463	10	US-09-737-178-25
17	118	6.2	996	118	US-09-815-242-5251
18	118	6.2	1009	10	US-09-815-242-12141
19	118	6.2	1115	9	US-10-153-273-2

Sequence 1, Appli
Sequence 123, App
Sequence 50, Appl
Sequence 2, Appli
Sequence 93, Appl
Sequence 41, Appl
Sequence 132, App
Sequence 5703, Ap
Sequence 12611, A
Sequence 107, App
Sequence 3, Appli
Sequence 46, Appl
Sequence 2, Appli
Sequence 4, Appli
Sequence 14, Appl
Sequence 32, Appl
Sequence 2, Appli
Sequence 57, Appl
Sequence 33, Appl
Sequence 12, Appl
Sequence 4, Appli
Sequence 218, App
Sequence 21, Appl
Sequence 1, Appli
Sequence 10, Appl
Sequence 2, Appli

20 116.5 6.1 290 10 US-09-808-885-1
21 114 5.9 452 10 US-09-737-178-123
22 107 5.6 1331 9 US-10-087-464-50
23 106.5 5.6 652 10 US-09-351-794A-2
24 106 5.5 521 9 US-09-820-843A-93
25 104.5 5.4 858 9 US-10-011-588-41
26 104.5 5.4 961 10 US-09-801-368-132
27 104 5.4 2025 10 US-09-815-242-5703
28 104 5.4 3158 10 US-09-815-242-12611
29 102.5 5.3 665 9 US-09-820-843A-107
30 100.5 5.2 394 9 US-09-978-756-3
31 100.5 5.2 428 9 US-10-087-464-46
32 99.5 5.2 383 9 US-10-098-514-2
33 99.5 5.2 383 9 US-10-098-514-4
34 99.5 5.2 402 9 US-10-098-514-14
35 99.5 5.2 497 9 US-09-820-843A-32
36 99 5.2 621 10 US-09-856-247A-2
37 99 5.2 2789 10 US-09-801-574-57
38 96.5 5.0 334 9 US-10-087-464-33
39 96.5 5.0 360 9 US-10-087-464-12
40 96.5 5.0 583 10 US-09-924-654-4
41 96.5 5.0 704 10 US-09-801-368-218
42 96.5 5.0 900 12 US-10-071-751-21
43 96.5 5.0 1295 10 US-09-726-949A-1
44 96.5 5.0 1639 9 US-10-087-464-10
45 96 5.0 376 9 US-09-978-756-2

ALIGNMENTS

RESULT 1

US-10-153-273-12

; Sequence 12, Application US/10153273

; Patent No. US20020169305A1

; GENERAL INFORMATION:

; APPLICANT: Sim, Kim L.

; Chitnls, Chetan

; Miller, Louis H.

; Peterson, David S.

; Su, Xin-zhaun

; Wellem, Thomas E.

; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

; NUMBER OF SEQUENCES: 37

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobbe Martens Olson & Bear

; STREET: 620 Newport Center Drive 16th Floor

; CITY: Newport Beach

; STATE: California

; COUNTRY: US

; ZIP: 92660

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/153,273

; FILING DATE: 21-May-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/210,288

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Fuller, Michael

; REGISTRATION NUMBER: 36,516

; REFERENCE/DOCKET NUMBER: NIH121.1FWDV1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 235-8550

; TELEFAX: (619) 235-0176

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 12;
US-10-153-273-12

Query Match 20.6%; Score 394.5; DB 9; Length 2710;
Best Local Similarity 26.2%; Pred. No. 7.7e-21;
Matches 122; Conservative 76; Mismatches 135; Indels 133; Gaps 20;

QY 1 YKEIQSY-----LSNDKFNVNINSE-YKQFYKELKETQYATNDTFLNLLNGEK 50 ~
DB 396 YDEIKKYGANGSGSRQKRDAGTTTNYDGYEKKFDELNKSRYRVDFLEKLSNEE 455
QY 51 YC-----KGLPGKDIITFTNSA-----DDKGFYRSEYQVCPDGVKCDGIKTYH 97
DB 456 ICTKVKDERGGTIDFKNVNSDSTSGASGNTVBSQGTFRYSKYCPCPYCGVK-----KVNN 511
QY 98 -KSDNDRVNN-----EDYKPPGVKPTNTITVLYSGNEQGDITQKLENECN-----143
DB 512 GGSSEWEKNGKCKGKLYEPKPDKEGTTITILKSGKHGDDIEKLNKFCDEKNGDTI 571
QY 144 ----SSTNYKDKNN-----QKWICYKDNINRCKLEQTEINNDNPK-----II-- 184
DB 572 NSGGSTGGSGGNGRQELYEKWK-YKEDVVKVGHDEDEEDYENVKNAGGLCILKN 630
QY 185 -----SFHNFELWYTLRLDTIKWNDKLTCTI-NNTTTHC-ID 221
DB 631 OKKNKEGGNTSEKPEDEIQTFFNFYYVAHMLKDSIHWKKLQRCIQNGNRKICGNN 690
QY 222 EGNRCLCFDRVVKQKEEWNISIKLFTKKNTIO-----QSYYSNINNLFEE 267
DB 691 KCNDCECFKRWITOKKDWGKIVQHF-KTONIKRGSGSDNTAEILPDPHDIVLYQINQOE 749
QY 268 GYF-----FKVMDKLDKDEAKWKELMENIKRKKNEFSNLENRD-----YLEN 310
DB 750 EFLGDSSEDAEESKSENSLDAEEA-----EELKHLREIESEDNNOEASVGGVTEQKN 803
QY 311 AIELLDHLKETATIC-----KDNNTNEACETSHNATNP 345
DB 804 IMDKLINYEKDEADLCLETHEDEEKEKGDGNECIEGENFRYP 849

RESULT 2

US-10-153-273-10
Sequence 10, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 700 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 10;
US-10-153-273-10

Query Match 19.8%; Score 380; DB 9; Length 700;
Best Local Similarity 28.7%; Pred. No. 1.9e-20;
Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

QY 1 YEKEI-----QSYLSNDNKFVNINSEYKQFYKELKETQYATNDTFL 43
DB 285 YETEISGGSGKSPKRTKRAARSSSSDDN-----GYESKFYKLEKVGQYQDVDFL 336
QY 44 NLLNEGKCKG---GLPGEKDIFTNSADKGFYRSEYQVCPDGVKCDGIKTYHKS 99
DB 337 KILNKEGICQKQVQNGEKADNVDFTEKYK-TFSRTEICEPCWGLEKGGPPWKVG 395
QY 100 DNDREVRVNEIDYKPPWGVKPTNITVLYSGNEQGDITQKLENFCSNSTNYKDKNNQKWCY 159
DB 396 DKTCGSATKIYDPK---NITDIPVLPDKSQONILKVKYKNCFKGAP-GGQIKKKWCY 451
QY 160 YKD-----ENINRCKLEQTEINNDNPKLIISFHNFFELWYTLRLDTIKWNDKLTCTINN 214
DB 452 YDEHRPSSKNNNNVEGTWDFKFTQKQTVKSYNVFFWDMVHMLHDSVFWKTELSKCIINN 511
QY 215 TTT-----HCIDECNRNCLCFDRVVKQKEEWNISIKLFTKKNIQSYYSNINNLFEGYF 270
DB 512 NTNGYTCRNNKCKTDGCFQKWVEKKQOEWMKIDHFGKQTDIVQQ-----KGLIVRSP 566
QY 271 FKVMDKLDKDEAKWKELMENIKRKKNEFSNLENNRDYLE-----NAIELLL 316
DB 567 YGVLDLV---LKGNNLLQNKIDVHGDITDKHKIKLLDEEDAVAVLGGKDNITIDKLL 622
QY 317 DHLKETATICKDNNTNEACE 336
DB 623 QHEKQAEQCKQK--QEECE 640

RESULT 3

US-10-153-273-8
Sequence 8, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

```

; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/153,273
; FILING DATE: 21-May-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,288
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8950
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-153-273-8

Query Match 12.0%; Score 231; DB 9; Length 921;
Best Local Similarity 24.1%; Pred. No. 31e-09;
Matches 91; Conservative 55; Mismatches 147; Indels 84; Gaps 17;

QY 36 YARNDFLNLNEGKCYCKGLPGEKDTFTNSADKGIYRSEYQVCP----DCGVKCD 91
DB 6 YTSIGLFLNSLKGPCCKKNDNAEDNIDF--GDEGTFKEADNCKPCSTVDC-KNCN 61
92 GIKYTHKSDNDRVRNVEDYKPPGWYKPTNITVLYSGNEQDITQKLE-----NF 141
DB 62 GDOTKGCNGSGKNGKNDY-----ITASDIENGSGNSIGNIDMVVSDKANGFNGLDA 114
142 CNSSTNYKKNQKWEYKDEINRCKLQO-NTEINNDNPKIISPHNFELWVYLLRD 200
DB 115 CGSANIEFGIRKEQWK-ACVGLDVCCLNGNGSIDKQKQIIIRALKRWVEYFLED 173
QY 201 TIKWNDKLKTCI--NTTTHCIDECNRCLCFDRWYKQKEEWNISIKLFF-TRKKNIQOS 257
DB 174 YNKINAKISHCTKKDNEST-CTNDPCNKCVCVEWIKRTEWKNIKKHKTQNEGDN 232
QY 258 YYSNNNL-----FEGYFFKVMKDKLDEAKWE-----LM 288
DB 233 MKSLVTDILGALQPSQDVNKAIRPCSGLTAFES--FCGLNGADNSEKKEGEDYDLVLCML 290
QY 289 ENIK-----RKNEFSNLEN-----NRDYLENAIELLDHLKETATC-----KD 328
DB 291 KNLKQIQEKKKHGTSVENGSKSTPLDNTTLEEEPIE--EENQVEAPNICPKQTVED 348
QY 329 NNTNEACETSHNATNP 345
DB 349 KKKEEBETCTPASVP 365

```

```

RESULT 4
US-09-820-843A-108
; Sequence 108, Application US/09820843A
; Publication No. US2003003963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PRO
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 108
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc_feature
; OTHER INFORMATION: gi|3845292
US-09-820-843A-108

Query Match 8.4%; Score 162; DB 9; Length 807;
Best Local Similarity 19.7%; Pred. No. 0.00035;
Matches 71; Conservative 55; Mismatches 118; Indels 116; Gaps 13;

QY 8 YLSNDKFNVINSEYKQFYEKLKEQYA--TNDTFLNLNEGKCYCKGLPGEKDTIF 64
DB 49 HMSNDS--NINKQENKKKKKTKHKNVNNININI-----84
QY 65 TNSADDRG-----IFYRSEYQVCPGCGVRCDCGIKTHKSDNDRVRNVEDYKPPW 115
DB 85 -HTTNDKNGDINKPEVIERDNIINIKNDTNLIDSSYNEEGENRNDNNNNNNNI 143
QY 116 GVPRTNITVLYSGNEQDITQKLENFCNSSTNYKDKNQKWEYKDEINRCKLEQNT 175
DB 144 NINNNI-----NNSCSNYYGLKKITLLKRNIDKDEGYN--ENITT 184
QY 176 INNDPKIISPHNFELWVYLLRDTIKWNDKLKTCINNTTTHCIDECNRCLCFDRWVK 235
DB 185 LNKNN-----LKNNNYNDNRNN--NNNNKNNNNNNNNCCSEKLE 226
QY 236 QKEEWNISIK-KLFT----KKKNIQOSYYSINNLFEGYFFKVMKDKLDEAKWKELMN 290
DB 227 QREKEYNKIRARISFNFNKQKQVQKTEQNNLNHTYLN-----NNIINN 270
QY 291 IKRKKNEFSNLEN-----NRDYLENAIELLDHLKETATICKDNNTNEACETSHN 340
DB 271 INNGDNOYAYINNFIYHNNSYNHYRONNI-----PICNINNHAPNIEKLNN 319

RESULT 5
US-09-924-154-16
; Sequence 16, Application US/09924154
; Patent No. US2002012741A1
; GENERAL INFORMATION:
; APPLICANT: Narum, David L.
; APPLICANT: Sim, Kim L.
; TITLE OF INVENTION: Anti-plasmodium Compositions and Methods of Use
; FILE REFERENCE: 05213-0465 43170-262105
; CURRENT APPLICATION NUMBER: US/09/924,154
; CURRENT FILING DATE: 2001-08-07
; PRIOR APPLICATION NUMBER: US 60/223,525
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 972
; TYPE: PRT
; ORGANISM: Mammalian

```

US-09-924-154-16

Query Match 8.4%; Score 161; DB 10; Length 972;
 Best Local Similarity 22.7%; Pred. No. 0.00052;
 Matches 77; Conservative 48; Mismatches 94; Indels 120; Gaps 21;
 QY 9 LSNDRKFN-----INSEYKOFYKEL--KETOYATNDTFLNLLNE-----48
 DB 105 ICONKNIENALYKIEDICNNTKVSIGELYCKE---KGNDKIWOQINEHIDKDPDVC 161
 QY 49 -----GKYCKGGLPGEKDIITFNSAD-----DKGIFVRSYC-----80
 DB 162 GPPRRQQLCLGNL--DKD-EFKVNDLKKFLNEILGIRDEGKFLIEKYRKNHMYLD 218
 QY 81 -QVCPDGVKCDGKIKYTHKSDNRERV--NNEDYKPPMGVKTPTNITVLYSNEQ---DI 134
 DB 219 ERAC-----KYLKSPDYKNIILGKDMWRDPSIKTENI---LKNFEGIKANI 265
 QY 135 TQKLENFCNS-----TNYKDN-NQWEC-----YKDNENTNRCKLEQNTNNDNPKII 184
 DB 266 VSMPSYADLSLDEPRKHWDQNKQKQWLEAISCEFYKGNHTGVCLMED-----DNDN---317
 QY 185 SPHNFELWYLLRDTIKWDLKATCINNTTHCIDR-----CNRNC 227
 DB 318 QYLHFWREKNDKFCIDKLNWDVIK-----EPICDKKVSPPSPNSPDVATVCKNSC 370
 QY 228 LCFDRWVQKEEWSIKKLFKKNIOQSYNS-INNL 265
 DB 371 TDYDKWILNKREY---KMQSSKYKRDRLFNVIQNL 405

RESULT 6

US-09-924-154-15
 ; Sequence 15, Application US/09924154
 ; Patent No. US20020127241A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Narum, David L.
 ; APPLICANT: Sim, Kim L.
 ; TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
 ; FILE REFERENCE: 05213-0465 43170-262105
 ; CURRENT APPLICATION NUMBER: US/09/924,154
 ; CURRENT FILING DATE: 2001-08-07
 ; PRIOR APPLICATION NUMBER: US 60/223,525
 ; PRIOR FILING DATE: 2000-08-07
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 15
 ; LENGTH: 1086
 ; TYPE: PRT
 ; ORGANISM: Mammalian

Query Match 7.9%; Score 151.5; DB 10; Length 1086;
 Best Local Similarity 18.4%; Pred. No. 0.003;
 Matches 87; Conservative 64; Mismatches 146; Indels 175; Gaps 20;
 QY 8 YLSNDKFNINSEYKOFYKELKETOYATNDTF--LMLNEG---KYCKGGLPGEKDI 62
 DB 212 YFAKDEPQIVRWLEWSKQFLDEKNYMLFTLRNTYNEMNIIHENNCKOYNKVVQNRKKEW 271
 QY 63 TETNSADKK-----GIF--YRS-----EYQVCPDGVKCDGIKT 96
 DB 272 TSLSENFKIFPERNVQIHSINIFKEYKENNVDIIFGLTNYEYNNFCKPELVSAAYN 331
 QY 97 HKSDNDRE-RV-----NNEDYKPPMGVKTPTN-----121
 DB 332 LKAPNAKSPRIYKSKHEESSVFGCKTKISKVKKKNCYSNNKVTKPEGVCGPPRRQQLC 391
 QY 122 ---ITVLYSGNEQ-----DITQKLEN-----FCNST-NYKD-----150
 DB 392 LGYIFLRDGNBGLKDHINKAANYEAMHLKEKENAGDGKICNAILGSIYADIGDIVRGL 451
 QY 151 -----KKNQKWCYKVDENINRCKLEQNTENN-----DNPXII 184

DB 452 DVWRDINTNKLSEKFOKIFMGSGNSRKQKQDNNERNKWKQKRNLIWSSVVKHPIPKKTC 511
 QY 185 SFHNFELWYLLRDTIKWDLK-----KTCINNTTHCIDENRNCCLCFDR 232
 DB 512 KRHNFEK-IPQFLRWLKEGDECEMGTEVYKOLEKICENKNCSE--KCKKNACSSYEK 568
 QY 233 WVQKEEWSIKKLFKKNIOQSYNSINNLFEYFFKVMKDKLDEAKWELMENIK 292
 DB 569 WIKERKNEYNLQSKFSDSKLNNK-----NNLYN-----KFEDSKAYLR 608
 QY 293 RKKNEFNLNDRYLENAIELLDHLKETATICKDNNTNEACETSHNATN 344
 DB 609 SESQCSNIEFN-----DETFTF--PNKYEACMVCEPNPSS 643

RESULT 7

US-09-924-154-13
 ; Sequence 13, Application US/09924154
 ; Patent No. US20020127241A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Narum, David L.
 ; APPLICANT: Sim, Kim L.
 ; TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
 ; FILE REFERENCE: 05213-0465 43170-262105
 ; CURRENT APPLICATION NUMBER: US/09/924,154
 ; CURRENT FILING DATE: 2001-08-07
 ; PRIOR APPLICATION NUMBER: US 60/223,525
 ; PRIOR FILING DATE: 2000-08-07
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 13
 ; LENGTH: 1421
 ; TYPE: PRT
 ; ORGANISM: Mammalian

Query Match 7.7%; Score 147; DB 10; Length 1421;
 Best Local Similarity 23.7%; Pred. No. 0.0089;
 Matches 68; Conservative 45; Mismatches 118; Indels 56; Gaps 16;
 QY 94 KYTHKSDNRERVANNEDYKPPMGVKTPTNITVLYSNEQ-GDIT-QKLENFCNSSTNYKDK 151
 DB 530 KYKNDKDEVCIIKNTF-----ADIRDIIGTDYNDLSNRKLVGKINTNSYVHR 581
 QY 152 NNQK-----WECYKDE-NINRCKLEQNTNNDN-PKILSFHNFELWYLLRD- 200
 DB 582 NKQNDKLFREDEWVKYKDKVNVISWVFKDKTVCKEDDIENIPQFFRWFSEWGDYCDQK 641
 QY 201 -----TIKNDKLTCTINNTTHCIDENRNCCLCFDRWVQKEEWSIKKLFKTK-KKNI 254
 DB 642 TKMIETLVECKEPCED-----DNCKRCKNSYKWKISKKKEYNKQAKQYQYQKGN 694
 QY 255 QOSYSNNINLNF-EGYFFKVMKDLK-----KDEAKWELMENIKRKN-----EFSNL 301
 DB 695 NYKYSEFSKPEVLYKLYSEKCSNLFNFEDEFK-EELHSDYKKNCTWCPEVKDVPISII 753
 QY 302 ENNRDYLENAI---ELLDLHLEKETATICKDNNTNEACETSHNATNP 345
 DB 754 RNNEOTSQEAPEESTEAHTETRT---DERKNO--EPANKDLKNP 795

RESULT 8

US-10-153-273-6
 ; Sequence 6, Application US/10153273
 ; Patent No. US20020169305A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sim, Kim L.
 ; APPLICANT: Chitnis, Chetan
 ; APPLICANT: Miller, Louis H.
 ; APPLICANT: Peterson, David S.
 ; APPLICANT: Su, Xin-zhaun
 ; APPLICANT: Wellens, Thomas E.

STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-153-273-4
Query Match 7.1%; Score 137; DB 9; Length 1435;
Best Local Similarity 20.5%; Pred. No. 0.05;
Matches 86; Conservative 57; Mismatches 149; Indels 128; Gaps 19;
QY 1 YEKEIQSYLSNDN-----KRVN-NINSEYKQFYERKLEKQYATNDTFLNLLNEGYCK- 53
Db 75 FNRKWSYGTDPDNDKXMSLNNKHNNEFMNNYQSFLTS-----SLIKQNKYVPI 126
QY 54 GGLPGKDJFTTNSADDKGIFYRSEYQVCPDQGVKQDGIKTHKSDNDRER-VNNEDYK 112
Db 127 NAVRVSRLSFLDSRINNG-RNTSSNNEVLSNCRKRGKMGKWDCKKKDRSNIVCIPDR 185
QY 113 PPGVVPPTNTVLYSGNEOD-----ITQKLEN-----FCNSSTN----- 147
Db 186 IQLCIVNLSITKTYKTMKDHPIEASKKESQLLLKKNDKNKYNKFCNDLKNSFLDYGHL 245
QY 148 -----YKDKNNOK-----WECYKD--E 163
Db 246 AMGNDMDPGYSTKAENKIQEVKGAHGEISEHKIKNFRKWNNEPREKLWEAMLSEHKN 305
QY 164 NINRCKLEQNTENNNDPKIISHPNFELWVYLLRDTKNDKDKLTCINNTHH-CIDE 222
Db 306 NINCKNIPOEEL-QITOWIKWHGEF-----LLERDNRSLKPKSK-CKNNTLYEACEKE 358
QY 223 CNRNCICFDWRVQKKEEWSIKKLTFTKKKNIQOYSYNNINNFEGYFFKVMKDKDEA 282
Db 359 CIDPCMRYDRIIRSKPEWHTLSKETQKVPKEN-----AENYLIKISE--NKDA 408
QY 283 KWKELENIKKRKNFNSNLENNRDYLENATELLDLHLKETATICK-----DNNTWEACE 336
Db 409 KVSLLNNCDAEYSKYCD-----CKHTTTLVKSVLNGNDNTIKERE 450
RESULT 11
US-09-402-100-4
; Sequence 4, Application US/09402100
; Patent No. US20010019834A1
; GENERAL INFORMATION:
; APPLICANT: Daewoong Pharmaceutical Co, LTD
; APPLICANT: Kim, Byung-O
; APPLICANT: Shin, Sung-Seup
; APPLICANT: Yu, Young-Hyo
; APPLICANT: Park, Myung-Hwan
; APPLICANT: Choi, Deok-Joon
; APPLICANT: Jung, Hyung-Jin
; TITLE OF INVENTION: Recombinant Microorganisms Expressing Antigenic Proteins of Hel
; FILE REFERENCE: 0136/0G140
; CURRENT APPLICATION NUMBER: US/09/402,100
; CURRENT FILING DATE: 1999-09-27
; EARLIER APPLICATION NUMBER: KR 97-11950
; EARLIER FILING DATE: 1997-03-31
; EARLIER APPLICATION NUMBER: KR 97-11951
; EARLIER FILING DATE: 1997-03-31
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 1338
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CagA/CTXA2B Chimeric protein
; US-09-402-100-4
Query Match 7.1%; Score 136; DB 10; Length 1338;
Best Local Similarity 19.0%; Pred. No. 0.055;
Matches 84; Conservative 69; Mismatches 139; Indels 150; Gaps 17;
QY 2 EKEIQS-----YLSNDNKFVNINSEYKQFYERKLE----- 33
Db 390 QEEIQNKIDFMEFLAQNNAKLDNLSEKEKEKFEITEKDFQKDSKAYLDALGNDRIAFVSK 449
QY 34 -----TOYATNDTFLNLLNEGYCKGGLPEKGDITFTNSADDKGIFYRSEYQVCP 84
Db 450 KDRHSALITEFGNGDLSYTLKDYKKADRALDREKNWTLQGSLLKHGVMF----- 500
QY 85 DCGYKCDGIKTHKSDNRE-----RVNNE 109
Db 501 ---VDYSNFKYTNASKNPNKGVNTGVSHLEVGFKVAIFNLPDLNLAITSFVRNLE 557
QY 110 DYKPPWGVKPTNITVL-----YSGNEQGDITQKLENF-----CNSSTNYK-----D 150
Db 558 DKLTTKGLSPQEAANKLIKDFLSSNK--ELVGKTLNFKAVADAANTGNYDEVKKAQKDL 615
QY 151 KNNQWCEYKID-----ENINRCKLEQNTENNNDPKIISHPNFELWVYLLRD--T 201
Db 616 KSLRRKREHEKEVEKLEKSGNKNKMEAKAQAQNSQKDEIFALIN-----KEANRDARA 669
QY 202 IKWNDKLTCTINNTHCHIDECNRNCLCFDRWVKOKEEWSIKKLTFTKKKNIQOYSYN 261
Db 670 IAYQNLAG-IRKELSDKLENVNNKLDKD--KSFDEFKNGKNKDSKAETLKALKGS 725
QY 262 INNLFEGYFFKVMKDKDEAKWELMENIKRKNFNSNLENNRDY-----LENAI 312
Db 726 VKDL-----GINPEWISKVENLNAALNEFKN-GKNKDFSKVTOAKSDLENSV 771
QY 313 ELLDLHLETATICKDNNTNEA 334
Db 772 KDVIINOKVTD---KVDNLNOA 790
RESULT 12
US-09-924-154-17
; Sequence 17, Application US/09924154
; Patent No. US20020127241A1
; GENERAL INFORMATION:
; APPLICANT: Narum, David L.
; APPLICANT: Sim, Kim L.
; TITLE OF INVENTION: Anti-plasmodium Compositions and Methods of Use
; FILE REFERENCE: 05213-0465 43170-262105
; CURRENT APPLICATION NUMBER: US/09/924,154
; CURRENT FILING DATE: 2001-08-07

; PRIOR APPLICATION NUMBER: US 60/223,525
 ; PRIOR FILING DATE: 2000-08-07
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 17
 ; LENGTH: 1501
 ; TYPE: PRT
 ; ORGANISM: Mammalian
 ; US-09-924-154-17

Query Match 6.9%; Score 133; DB 10; Length 1501;
 Best Local Similarity 22.1%; Pred. No. 0.1;
 Matches 79; Conservative 46; Mismatches 153; Indels 80; Gaps 15;

QY 7 SYLSNDKFNINSEYKQFYEKLK---TOYATNDTFLNLLNEGYKGGPLGGEKIDT 63
 DB 474 NYLLDDIY-NVHNSQLLIEIMASKQEGKLLMKHGTILDONACKYIINDSYVDYKDI 532
 QY 64 FTNSADDKGIFYRSEYQVCPDCGVKCDGKIYTHKSDNDRERNEDYKPPGWVKPNTIT 123
 533 IGNDL-----WNNDSIKVON-----NLN 551
 QY 124 VLYSGNEQGDY-TQKLENFCSNSTN--YKDKNNQKWE---CYKNDENINRCKLEQNTIN 177
 DB 552 LIPERFEGYKVGKRLFKTKELKNVWMLNRKNKVESMRGIDEDVDORRKTCEIDELE 611
 QY 178 NDNPKIISFNFELWYLLRTDIKNDKL--KTCINNTTTHCIDB-CNRNCLCFDRW- 233
 DB 612 N---MPQFTRFWSQAHFFCKEYKELWELKNDKCTGNGKSLCQDKTCQNVCTNNMYT 667
 QY 234 -VKQKEEWSIK----KLFT--KKNIQOQSYVSNINNLFEYFFKVMKLDKDEAKWK 285
 DB 668 YTRKLAIEIOSYKDKRKLFLSLAKORNVTFLEKAKNSINDFTKIFDQDK---LFK 724
 QY 286 ELMENIKRKNFNSNENRNDYLENAJELLHLKETATICKDNNTNEACETSHNAT 343
 DB 725 ERCSMDTQVLEVNKE-----MLSDNSSEDATDISEKNGEELYVNHNSVS 772

RESULT 13

US-09-924-154-14
 ; Sequence 14, Application US/09924154
 ; Patent No. US20020127241A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Narum, David L.
 ; APPLICANT: Sim, Kim L.
 ; TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
 ; FILE REFERENCE: 05213-0465 43170-262105
 ; CURRENT APPLICATION NUMBER: US/09/924,154
 ; CURRENT FILING DATE: 2001-08-07
 ; PRIOR APPLICATION NUMBER: US 60/223,525
 ; PRIOR FILING DATE: 2000-08-07
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 14
 ; LENGTH: 1143
 ; TYPE: PRT
 ; ORGANISM: Mammalian
 ; US-09-924-154-14

Query Match 6.7%; Score 128; DB 10; Length 1143;
 Best Local Similarity 19.2%; Pred. No. 0.18;
 Matches 84; Conservative 70; Mismatches 151; Indels 132; Gaps 21;

QY 3 KEIQSYLSNDN-----KFNINSEYKQFYEKLKETOYATNDTFLNLLNEGYKGG 55
 DB 74 KLINSFVENKSVKKRSLSPFNNTKS-----YDIIPSYSYRNDKF-NLSSENEEDNSGN 127
 QY 56 LPGEKIDTFNSA-----DDKGYFYRSEYQVCPDCGVKCDGKIYTHKSDNDRERY-- 106
 DB 128 TNSN---NFANTSISIGKDNKQYTFQKRHLFA-CGIRKRSIKWICRENSKITVCP 183
 QY 107 -----NNEDYKPPGWVK-PTNTVLYSGNEQGD-----ITQKLENF 141

DB 184 DRKIQLCIANFLNSRLETMEKREIFLISVNTAKLLYNKNEGKDSIFCNELENSFSDF 243
 QY 142 CNS-----STNYKDKNNQKWEYKDENI-----NRCKLEQNTIN 177
 DB 244 RNSFIGDDMDFGGNTDRVKGVIYKKEKNEVKNLNNIKKEWEKKNANLWNHIVN 303
 QY 178 N-----DNPKIISFNFELWYLLRTDIKNDKL-----KTCI 212
 DB 304 HKGNISKECAIIPAEPOI-----NLWK-----EWNENFLMEKKRLFLNKDKCV 349
 QY 213 NNTTTH-CIDECNRNCLCFDRWVKQKEEWSIKKLTFTKKKNIQOQSYVSNINNLFEYGF 271
 DB 350 ENKYEACFCGCGRLPCSSYTSFPMKSKTQMEVLTNLY-KKNSGVDKNNFLNLFKANK 408
 QY 272 KVMDKLDKDEAKWKELME-----NIKKRKNFNSNENRNDYLENAJELLHLKETATI 325
 DB 409 NLDLDPFKEKEVDLDCRYTATIK---SFLNGPAKND-VDIASQINNVNDRGRGFCN 463
 QY 326 CKDNNTNEACETSHNAT 342
 DB 464 YKSN-----EKSWNCT 475

RESULT 14

US-09-820-843A-73
 ; Sequence 73, Application US/09820843A
 ; Publication No. US20030039963A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Council of Scientific and Industrial Research
 ; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PRO
 ; FILE REFERENCE: 063915
 ; CURRENT APPLICATION NUMBER: US/09/820,843A
 ; CURRENT FILING DATE: 2001-03-30
 ; NUMBER OF SEQ ID NOS: 118
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 73
 ; LENGTH: 1805
 ; TYPE: PRT
 ; ORGANISM: M. genitalium
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: gi|1045905
 ; US-09-820-843A-73

Query Match 6.4%; Score 122; DB 9; Length 1805;
 Best Local Similarity 20.5%; Pred. No. 0.85;
 Matches 81; Conservative 57; Mismatches 121; Indels 136; Gaps 18;

QY 6 QSYLSNDKFNINSEYKQFYEKLKETOYATNDTFLNLLNEGYKGGPLGGEKIDTFT 65
 DB 1234 QTYLANKN-----AEY-----SQOOLQKYTNLID-----LKENLERTKD--- 1269
 QY 66 NSADDKGIFYRSEYCOV---CPDCGVKCDGKIYTHKSDNDRERNEDYKPPGWVKPTNI 122
 DB 1270 -QDKK---HRSIFARLTKFANDRFEKKQLLKAQRIYVDDKNRLKENER----- 1315
 QY 123 TVLYSGNEQGDITOKLENFCSNSTNYKDKNNQK-----WECYKNDENINRCKLE-- 171
 DB 1316 NLHLSNE---TERKRAVLEDOISYFEKQKQATDAILASHKEVKKKEGELQKLLVELE 1371
 QY 172 -QNTENNDNPKIISFNFELWYLLRTDIKNDKLK-----TCINNTTTHC 219
 DB 1372 TRKTKLANDFAKFSRQREEF-----NORLLELOKLTQTOTNSNFKTKA 1418
 QY 220 IDECNRLCLCFDRWVKQKEEWSIKKLTFTKKKNIQOQSYVSNINNLFEYKFKVMDKLDK 279
 DB 1419 IQEITENS-----YKRGMEELNFKQKEFDKN-----SRLYE-YFRKMRDEIER 1460
 QY 280 DEAKWKELMENIKRKNFNS-----NLENNR-DYLENAJELLHL----- 319
 DB 1461 KESQVKLVLETKQKANKLLEAQANKLEKNTIDFKELKAFKDKVDQDIDSTNKORKE 1520

QY 320 -----KETATICKDNNNEACET 337
Db 1521 LNELNENKLLQSLIERERAINSKDLSLNKKIET 1555

RESULT 15
US-09-286-488-25
; Sequence 25, Application US/09286488
; Patent No. US20020169136A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C3
; CURRENT APPLICATION NUMBER: US/09/286,488
; CURRENT FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Babesia microti
US-09-286-488-25

Query Match 6.3%; Score 120; DB 9; Length 463;
Best Local Similarity 22.4%; Pred. NO. 0.24;
Matches 92; Conservative 43; Mismatches 146; Indels 130; Gaps 20;

QY 2 EXEIQYLSNDNKFVNNSEYKQYKLEKQYATNDTEL-----NLLNE 48
Db 84 DKSQDKLSHPNKIDKIKIDSIIEFDNAKLPTGSVDISITCKHNNPVLIRFSLIE 143
QY 49 GKCYK-----GGLPGEKIDITFTNSADDKGI-FYRSEYQVCQPCD--GVKCDGI 93
Db 144 GSCYFYFLLNNDTNKNNHKLKDYKTNEHTDNGINGINYKIDYSESTEPTTCTCF 203
QY 94 -KYTHKSDNDRRVNNEDYKPPWGVKPTNITVLYSGNEQGDITQKLEFCN-----SS 145
Db 204 RKNHKS---ERKELENYK-----YEGTELARI-----HCNKGKCVKLG 240
QY 146 TNYKDNKNQKWCYKDEINRCKLEQNTNNDNPKIISFHNFFELWYTLIRDTIKWN 205
Db 241 IKIKDN---LEIYVQ-----LMSVNPVNFNDPTSIN-----LPTVSTTNDIT-N 284
QY 206 DLKTCIN-NTTTHC-----IDECNRNCLC-----FDRWV 234
Db 285 KYGTIINANIVECEPEDEPLTIGFRTIDKSQLSHPNKIDKIKFFDYIIEFDDV 344
QY 235 KOREEWNISIKLETKKN-----IQOSYV-----SNINLFEGYFPKVMKLD 278
Db 345 KLPTIGTVNIYIYTCBHNPNVLFVIFVLSIESYFYFYSMMNNTKNNHKLKDYKRFK 404
QY 279 KDEAKWELMENIKRK-----KNFSNLNRRDYL--ENAIELLDHLKE 321
Db 405 KYTKNGINCYEYVLRKCSSYTRKNVEYHKLARIHCNEEKCVNVKVDNIEK 455

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:33:00 ; Search time 58.8889 Seconds
(without alignments)
780.648 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747

Perfect score: 345

Sequence: 1 YEKEIQSYLSNDNKFVNIN.....CKDNTNEACETSHNATNP 345

Scoring table:

OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

d size: 9

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: A_Geneseq_101002:*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	345	100.0	3542	22	AAB62142 P. falciparum FCR3

ALIGNMENTS

RESULT 1:
AAB62142
ID AAB62142 standard; Protein; 3542 AA.

XX

AC AAB62142;

XX 29-MAY-2001 (first entry)

XX

DE P. falciparum FCR3.varCSA protein.

XX

KW FCR3.varCSA protein; chondroitin sulfate A (CSA); var gene: PPEMP1;
KW erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
KW malaria; protozoacide.

XX

OS Plasmodium falciparum.

XX

PN WO200116326-A2.

XX

PD 08-MAR-2001.

XX

PF 01-SEP-2000; 2000WO-US24195.

XX

PR 01-SEP-1999; 99US-0152023.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX

PI Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
PI Gysin J, Pouvelle B, Fujii N, Smith J;

XX

XX WPI; 2001-235109/24.

DR N-PSDB; AAF57301.

XX

PT Novel FCR3.varCSA protein, useful for modulating parasitized red blood
PT cell binding, sequestration and onset of maternal malaria -
XX Claim 12; Page 63-71; 78pp; English.

XX The invention relates to a P. falciparum FCR3.varCSA protein, that is
CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
CC corresponding P.falciparum erythrocyte membrane protein 1 (PPEMP1)
CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
CC protein and the encoding gene are useful for treating and preventing
CC maternal malaria in a patient identified at a risk for contracting
CC the maternal malaria or in a patient afflicted with maternal malaria. The
CC present sequence represents the P. falciparum FCR3.varCSA protein.

XX Sequence 3542 AA;

Query Match 100.0%; Score 345; DB 22; Length 3542;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEKEIQSYLSNDNKFVNINSEYKQYKELKETOYATNDTFLNLLNEGKYGKGLPGEK 60

Db 403 YEKEIQSYLSNDNKFVNINSEYKQYKELKETOYATNDTFLNLLNEGKYGKGLPGEK 462

QY 61 DIFTTSADDKGIFIRSEYQVCPDCGVKCDGKIKYTHKSDNDRVNNEDYKPPWGVKPT 120

Db 463 DIFTTSADDKGIFIRSEYQVCPDCGVKCDGKIKYTHKSDNDRVNNEDYKPPWGVKPT 522

QY 121 NITVLYSGNEQGDITQKLENFCNSSTNYKDNKQKWEYKDKENINRCKLEQNTENNNDN 180

Db 523 NITVLYSGNEQGDITQKLENFCNSSTNYKDNKQKWEYKDKENINRCKLEQNTENNNDN 582

QY 181 PKLIISFHNFFELWYLLRDTIKWNDKLTICINNTTHCIDEKNRNCICLDFRWVKQKEE 240

Db 583 PKLIISFHNFFELWYLLRDTIKWNDKLTICINNTTHCIDEKNRNCICLDFRWVKQKEE 642

QY 241 WNSIKKLTFTKKKNIQSSYYSNNINLFGYFFKVMYDKLDEAKWKELMENIKRKNFESN 300

Db 643 WNSIKKLTFTKKKNIQSSYYSNNINLFGYFFKVMYDKLDEAKWKELMENIKRKNFESN 702

QY 301 LENNRDYLENAIELLLDHLKETATICKDNTNEACETSHNATNP 345

Db 703 LENNRDYLENAIELLLDHLKETATICKDNTNEACETSHNATNP 747

Search completed: April 28, 2003, 10:49:42
Job time : 59.3889 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:48:21 ; Search time 31.1111 Seconds
(without alignments)
1066.062 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747
Perfect score: 345
Sequence: 1 YEKEIQSYLSNDNKFVNIN.....CKDNTNEACETSHNATTNP 345

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

d size : 9

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
---------------	-------	----------------	--------	----	-------------

No matches found

rch completed: April 28, 2003, 10:52:27
time : 31.1111 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:35:40 ; Search time 17.7778 Seconds
(without alignments)
804.900 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747
Perfect score: 345
Sequence: 1 YEKEIQSYLSNDNKFVNIN.....CKDNTNEACETSHNATTNP 345

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 112892 seqs, 41476328 residues

d size : 9
Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description

No matches found

Search completed: April 28, 2003, 10:50:21
Job time : 18.7778 secs

THIS PAGE BLANK (usps)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:45:26 ; Search time 31.6667 Seconds
(without alignments)
2244.829 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747

Perfect score: 345

Sequence: 1 YEKEIQSYLSNDKFNIN.....CKDNTNEACETSHNATNP 345

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

d size : 9

Total number of hits satisfying chosen parameters: 14

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- 1: SP_TREMBL_21.*
- 2: sp_archaea.*
- 3: sp_bacteria.*
- 4: sp_fungi.*
- 5: sp_human.*
- 6: sp_invertebrate.*
- 7: sp_mammal.*
- 8: sp_mhc.*
- 9: sp_organelle.*
- 10: sp_phage.*
- 11: sp_plant.*
- 12: sp_rodent.*
- 13: sp_virus.*
- 14: sp_vertebrate.*
- 15: sp_unclassified.*
- 16: sp_rvirus.*
- 17: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	345	100.0	3542	5	Q9U5M2	Q9U5M2 plasmodium
2	240	69.6	461	5	Q8T6K3	Q8T6K3 plasmodium
3	240	69.6	465	5	Q8T6K4	Q8T6K4 plasmodium
4	240	69.6	492	5	Q8T6K9	Q8T6K9 plasmodium
5	238	69.0	494	5	Q8T6K7	Q8T6K7 plasmodium
6	233	67.5	510	5	Q8T6L0	Q8T6L0 plasmodium
7	67	19.4	492	5	Q8T6K6	Q8T6K6 plasmodium
8	67	19.4	1615	5	Q8T325	Q8T325 plasmodium
9	49	14.2	3287	5	Q8T326	Q8T326 plasmodium
10	41	11.9	460	5	Q8T6K1	Q8T6K1 plasmodium
11	41	11.9	494	5	Q8T6K0	Q8T6K0 plasmodium
12	39	11.3	427	5	Q8T6K2	Q8T6K2 plasmodium
13	39	11.3	496	5	Q8T6K8	Q8T6K8 plasmodium
14	34	9.9	494	5	Q8T6K5	Q8T6K5 plasmodium

ALIGNMENTS

RESULT 1

Q9U5M2 ID Q9U5M2 PRELIMINARY; PRT; 3542 AA.
AC Q9U5M2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE FC3 CSA ligand (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FCR3;
RX MEDLINE=20006305; PubMed=10535993;
RA Buffet P., Gamain B., Scheideig C., Baruch B., Oishi S., Fujii N.,
RA Fusai T., Parzy D., Miller L.H., Gysin J., Scherf A.;
RT "Plasmodium falciparum domain mediating adhesion to chondroitin
sulfate A: A receptor for human placental infection.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:12743-12748(1999).
DR EMBL: AJ133811; CAB59840.1; -;
DR InterPro: IPR001219; Neurotoxin.
DR InterPro: IPR004258; PFEMP.
DR Pfam: PF03011; PFEMP; 1.
DR PRINTS: PR00284; TOXIN.
FT NON_TER 3542 3542
SQ SEQUENCE 3542 AA; 413089 MW; 970D85EE88DA2EC2 CRC64;
Query Match 100.0%; Score 345; DB 5; Length 3542;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 YEKEIQSYLSNDKFNINSEYKQFYEKLKETATNDTFLNLLNEGYCKGGLPGEK 60
|||||
403 YEKEIQSYLSNDKFNINSEYKQFYEKLKETATNDTFLNLLNEGYCKGGLPGEK 462
QY 61 DITFTNSADDKGFYRSEYQVCPDCGKDGKIKYTHKSDNDRVNNEDYKPPWGVKPT 120
|||||
463 DITFTNSADDKGFYRSEYQVCPDCGKDGKIKYTHKSDNDRVNNEDYKPPWGVKPT 522
QY 121 NITVLYSGNEQGDITQKLENFCNSSTNYKDKNNQKWEYKDKENINRCKLEQNTENNND 180
|||||
523 NITVLYSGNEQGDITQKLENFCNSSTNYKDKNNQKWEYKDKENINRCKLEQNTENNND 582
QY 181 PKIISPHNFELWVTVLLRDITKWNKDKTKTCINNTTTHCIDECNRNCLCFDRVVKKEE 240
|||||
583 PKIISPHNFELWVTVLLRDITKWNKDKTKTCINNTTTHCIDECNRNCLCFDRVVKKEE 642
QY 241 WNSIKKLFKKNIQOYSYNNINLFEYFFKVMYDLDKDEAKWELMENIKRKKNEFSN 300
|||||
643 WNSIKKLFKKNIQOYSYNNINLFEYFFKVMYDLDKDEAKWELMENIKRKKNEFSN 702
QY 301 LENNRDYLENAIELLDHLKETATICKDNTNEACETSHNATNP 345
|||||
703 LENNRDYLENAIELLDHLKETATICKDNTNEACETSHNATNP 747

RESULT 2

Q8T6K3 ID Q8T6K3 PRELIMINARY; PRT; 461 AA.
AC Q8T6K3;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;

```

RN RP SEQUENCE FROM N.A.
RC STRAIN-M58;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
RL J. Infect. Dis. 0:0-0(2002).";
DR EMBL; AF458893; AAL84276.1; -.
FT NON_TER 1 461
FT NON_TER 461 461
SQ SEQUENCE 461 AA; 54946 MW; F4C8863D50DA50EF CRC64;

Query Match 69.6%; Score 240; DB 5; Length 461;
Best Local Similarity 100.0%; Pred. No. 9.7e-237;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 73
DB 180 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 239
74 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 133
240 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 299
QY 134 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 193
DB 300 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 359
QY 194 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 253
DB 360 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 419

RESULT 3
QY 14 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 73
DB 180 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 239
74 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 133
240 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 299
QY 134 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 193
DB 300 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 359
QY 194 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 253
DB 360 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 419

Query Match 69.6%; Score 240; DB 5; Length 465;
Best Local Similarity 100.0%; Pred. No. 9.8e-237;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 73
DB 183 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 242
74 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 133
243 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 302
QY 134 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 193
DB 303 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 362

Query Match 69.6%; Score 240; DB 5; Length 465;
Best Local Similarity 100.0%; Pred. No. 9.8e-237;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 73
DB 183 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 242
74 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 133
243 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 302
QY 134 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 193
DB 303 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 362
```

```

QY 194 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 253
DB 363 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 422

RESULT 4
QY 194 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 253
DB 363 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 422

Query Match 69.6%; Score 240; DB 5; Length 492;
Best Local Similarity 100.0%; Pred. No. 1e-236;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 73
DB 182 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 241
74 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 133
242 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 301
QY 134 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 193
DB 302 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 361
QY 194 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 253
DB 362 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 421

Query Match 69.6%; Score 240; DB 5; Length 492;
Best Local Similarity 100.0%; Pred. No. 1e-236;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 73
DB 182 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 241
74 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 133
242 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 301
QY 134 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 193
DB 302 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 361
QY 194 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 253
DB 362 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 421

Query Match 69.6%; Score 240; DB 5; Length 494;
Best Local Similarity 100.0%; Pred. No. 1e-236;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 73
DB 182 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 241
74 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 133
242 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 301
QY 134 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 193
DB 302 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 361
QY 194 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 253
DB 362 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 421

Query Match 69.6%; Score 240; DB 5; Length 494;
Best Local Similarity 100.0%; Pred. No. 1e-236;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 73
DB 182 KVVNNINSEYKQFYEKLKQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 241
74 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 133
242 FYRSEYQVCPDQCGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 301
QY 134 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 193
DB 302 ITQKLENFCNSSTNYKDKNNQWECYKQDENINRCKLEQNTNINNDNPKIISFHNFFELW 361
QY 194 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 253
DB 362 VTYLLRDTIKWNDKLTCTINNTTTHCIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 421
```

```
Query Match          69.0%; Score 238; DB 5; Length 494;
Best Local Similarity 100.0%; Pred. No. 1.1e-234;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KFNINSEYKQFYKELKETOYATNDTFLNLLNEGKCKGGLPGKDIPTNSADDKGI 73
DB 182 KFNINSEYKQFYKELKETOYATNDTFLNLLNEGKCKGGLPGKDIPTNSADDKGI 241

QY 74 FYRSEYQVCPGKVKCDGKIKYTHKSDNDRVRNNEDYKPPMGVKPTNTIVLYSGNEQGD 133
DB 242 FYRSEYQVCPGKVKCDGKIKYTHKSDNDRVRNNEDYKPPMGVKPTNTIVLYSGNEQGD 301

QY 134 ITQKLEFNCNSTNYKDKNNQWECYKDEINRCKLEQNTINNDNPKIISFHNFFELW 193
DB 302 ITQKLEFNCNSTNYKDKNNQWECYKDEINRCKLEQNTINNDNPKIISFHNFFELW 361

QY 194 VTYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSIKKLTFFK 251
DB 362 VTYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSIKKLTFFK 419

RESULT 6
Q8T6L0
ID Q8T6L0 PRELIMINARY; PRT; 510 AA.
AC Q8T6L0;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implicated in malaria in pregnancy.";
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL; AF45886; AAL84269.1;
FT NON_TER 1
FT NON_SEQUENCE 510 AA; 60713 MW; E3CF55DC04601099 CRC64;
SQ SEQUENCE 510 AA; 60713 MW; E3CF55DC04601099 CRC64;

Query Match          67.5%; Score 233; DB 5; Length 510;
Best Local Similarity 100.0%; Pred. No. 1.5e-229;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KFNINSEYKQFYKELKETOYATNDTFLNLLNEGKCKGGLPGKDIPTNSADDKGI 73
DB 184 KFNINSEYKQFYKELKETOYATNDTFLNLLNEGKCKGGLPGKDIPTNSADDKGI 243

QY 74 FYRSEYQVCPGKVKCDGKIKYTHKSDNDRVRNNEDYKPPMGVKPTNTIVLYSGNEQGD 133
DB 244 FYRSEYQVCPGKVKCDGKIKYTHKSDNDRVRNNEDYKPPMGVKPTNTIVLYSGNEQGD 303

QY 134 ITQKLEFNCNSTNYKDKNNQWECYKDEINRCKLEQNTINNDNPKIISFHNFFELW 193
DB 304 ITQKLEFNCNSTNYKDKNNQWECYKDEINRCKLEQNTINNDNPKIISFHNFFELW 363

QY 194 VTYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSIKK 246
DB 364 VTYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSIKK 416

RESULT 7
Q8T6K6
ID Q8T6K6 PRELIMINARY; PRT; 492 AA.
AC Q8T6K6;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
```

```
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=M48;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implicated in malaria in pregnancy.";
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL; AF45886; AAL84273.1;
FT NON_TER 1
FT NON_SEQUENCE 492 AA; 58746 MW; 49D385FF0914B34A CRC64;
SQ SEQUENCE 492 AA; 58746 MW; 49D385FF0914B34A CRC64;

Query Match          19.4%; Score 67; DB 5; Length 492;
Best Local Similarity 100.0%; Pred. No. 8e-60;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 SFHNFFELWTVYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSI 244
DB 354 SFHNFFELWTVYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSI 413

QY 245 KKLFTKK 251
DB 414 KKLFTKK 420

RESULT 8
Q8T325
ID Q8T325 PRELIMINARY; PRT; 1615 AA.
AC Q8T325;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE PEMP1 (Fragment).
GN TM284VAR3.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TM284;
RX MEDLINE=21927235; PubMed=11930336;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implicated in malaria in pregnancy.";
RL J. Infect. Dis. 185:1207-1211(2002).
DR EMBL; AJ420412; CAD20868.1;
FT NON_TER 1615
FT NON_SEQUENCE 1615 AA; 188639 MW; FCA896C00B86DEA6 CRC64;
SQ SEQUENCE 1615 AA; 188639 MW; FCA896C00B86DEA6 CRC64;

Query Match          19.4%; Score 67; DB 5; Length 1615;
Best Local Similarity 100.0%; Pred. No. 2.1e-59;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 SFHNFFELWTVYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSI 244
DB 577 SFHNFFELWTVYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSI 636

QY 245 KKLFTKK 251
DB 637 KKLFTKK 643

RESULT 9
Q8T326
ID Q8T326 PRELIMINARY; PRT; 3287 AA.
AC Q8T326;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
```

```

DE PfEXPl (Fragment).
GN TM180VAR2.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TM180.
RX MEDLINE=21927235; PubMed=11930336;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implicated in malaria in pregnancy.";
RL J. Infect. Dis. 185:1207-1211(2002).
DR EMBL; AJ420411; CAD20867.1; -.
FT NON_TER 3287 3287
SQ SEQUENCE 3287 AA; 383550 MW; 58F8F866FC244536 CRC64;

Query Match 14.2%; Score 49; DB 5; Length 3287;
Best Local Similarity 100.0%; Pred. No. 9.8e-41;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

297 EFSNLENNRDYLENAIELLDHLKETATICKDNNTEACETSHNATNP 345
|||||
689 EFSNLENNRDYLENAIELLDHLKETATICKDNNTEACETSHNATNP 737

RESULT 10
Q8T6K1 PRELIMINARY; PRT; 460 AA.
ID Q8T6K1
AC Q8T6K1;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VAN002;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implicated in malaria in pregnancy.";
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL; AF458895; AAL84278.1; -.
FT NON_TER 460 460
SQ SEQUENCE 460 AA; 55157 MW; 06FC87959F5C7B9E CRC64;

Query Match 11.9%; Score 41; DB 5; Length 460;
Best Local Similarity 100.0%; Pred. No. 2.9e-33;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CINNTHTCIDECNRNCLCFDRWVKQKEEWSIKKLFYKK 251
|||||
DB 375 CINNTHTCIDECNRNCLCFDRWVKQKEEWSIKKLFYKK 415

RESULT 11
Q8T6K0 PRELIMINARY; PRT; 494 AA.
ID Q8T6K0
AC Q8T6K0;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-VAN434;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implicated in malaria in pregnancy.";
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL; AF458896; AAL84279.1; -.
FT NON_TER 494 494
SQ SEQUENCE 494 AA; 59240 MW; 739BE121652A367B CRC64;

Query Match 11.9%; Score 41; DB 5; Length 494;
Best Local Similarity 100.0%; Pred. No. 3.1e-33;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CINNTHTCIDECNRNCLCFDRWVKQKEEWSIKKLFYKK 251
|||||
DB 381 CINNTHTCIDECNRNCLCFDRWVKQKEEWSIKKLFYKK 421

RESULT 12
Q8T6K2 PRELIMINARY; PRT; 427 AA.
ID Q8T6K2
AC Q8T6K2;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TM4C8;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implicated in malaria in pregnancy.";
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL; AF458894; AAL84277.1; -.
FT NON_TER 427 427
SQ SEQUENCE 427 AA; 50969 MW; 6AF7922CB988F20 CRC64;

Query Match 11.3%; Score 39; DB 5; Length 427;
Best Local Similarity 100.0%; Pred. No. 3e-31;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 SFHNFELWVYLLRDTIKWNDKLKTCINNTTHCIDEK 223
|||||
DB 354 SFHNFELWVYLLRDTIKWNDKLKTCINNTTHCIDEK 392

RESULT 13
Q8T6K8 PRELIMINARY; PRT; 496 AA.
ID Q8T6K8
AC Q8T6K8;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implicated in malaria in pregnancy.";
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL; AF458888; AAL84271.1; -.
FT NON_TER 496 496
SQ SEQUENCE 496 AA; 59240 MW; 739BE121652A367B CRC64;

```

SQ SEQUENCE 496 AA; 59269 MW; A9569625061F9411 CRC64;

Query Match 11.3%; Score 39; DB 5; Length 496;
Best Local Similarity 100.0%; Pred. No. 3.4e-31;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 185 SPHNFELWVYLLRDTIKWNDKLTCTINNTTTHCIDE 223

Db 355 SPHNFELWVYLLRDTIKWNDKLTCTINNTTTHCIDE 393

RESULT 14

O8T6K5

ID O8T6K5 PRELIMINARY; PRT; 494 AA.

AC O8T6K5;

DT 01-JUN-2002 (TREMBlrel. 21, Created)

DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Erythrocyte membrane protein 1 (Fragment).

VAR.

Plasmodium falciparum.

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K117;

RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;

RT Identification of a conserved Plasmodium falciparum var gene

RT implicated in malaria in pregnancy.;

RL J. Infect. Dis. 0:0-0(2002).

DR EMBL: AF458891; AAL84274.1;

FT NON_TER 1

FT NON_TER 494

SQ SEQUENCE 494 AA; 58870 MW; 56A34264A9285321 CRC64;

Query Match 9.9%; Score 34; DB 5; Length 494;

Best Local Similarity 100.0%; Pred. No. 4.5e-26;

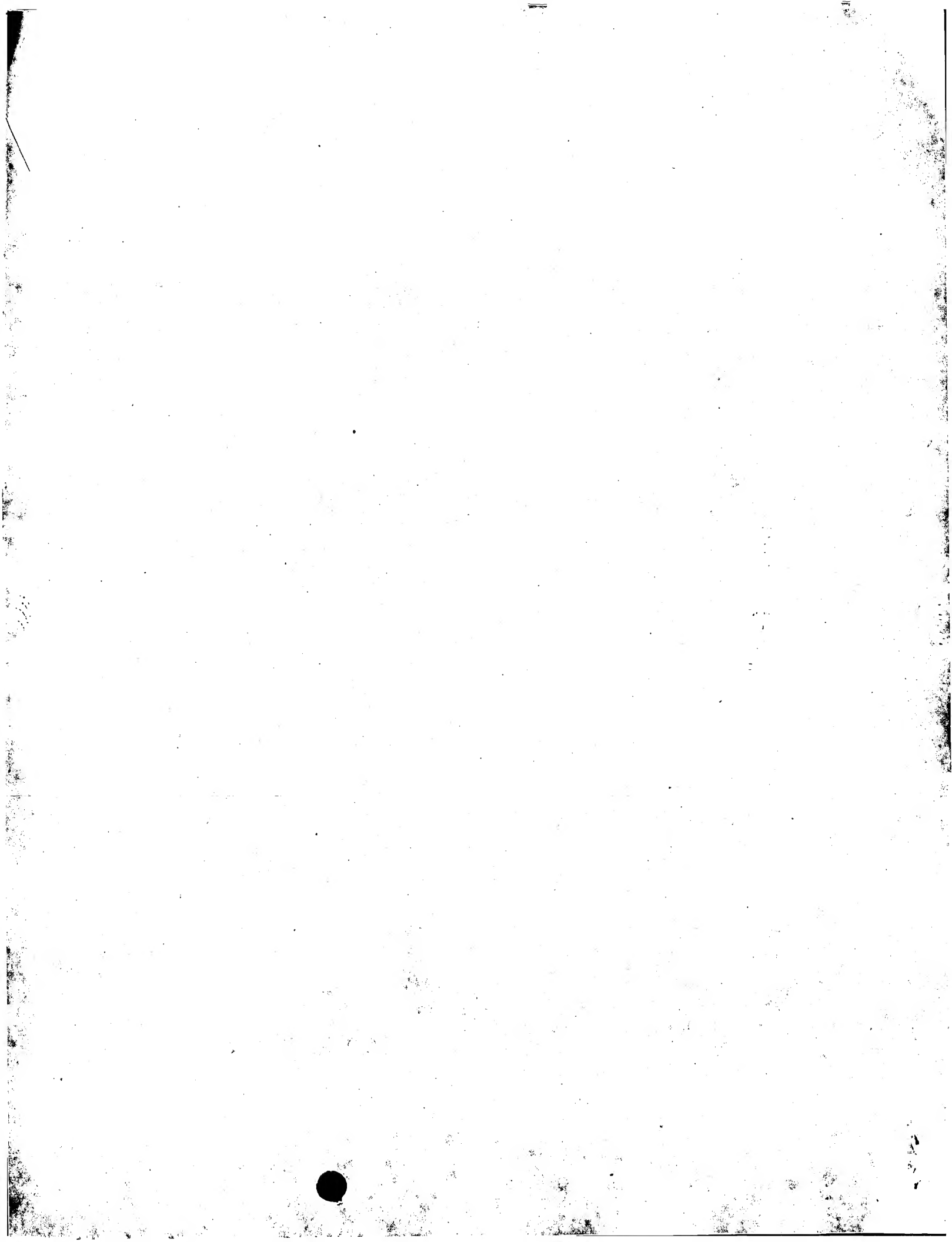
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 144 SSTNYKDKNNQKWCYKDNINRCKLEQNTIN 177

Db 315 SSTNYKDKNNQKWCYKDNINRCKLEQNTIN 348

Search completed: April 28, 2003, 10:51:26

Job time : 33.6667 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:49:21 ; Search time 21.1111 Seconds
(without alignments)
480.832 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747

Perfect score: 345

Sequence: 1 YEKEIQSYLSNDKFNIN.....CKDNTNEACEISHNATNP 345

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

d size : 9

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	DB	ID	Description
---------------	----------------	-------	--------	----	----	-------------

No matches found

Search completed: April 28, 2003, 10:53:12
Job time : 21.1111 secs



GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:51:36 ; Search time 16.6687 Seconds
(without alignments)
1658.687 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747
Perfect score: 345
Sequence: 1 YEKEIQSYLSNDKFNIN.....CKDNTNEACETSHNATNP 345

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 301932 seqs, 80129803 residues

d size : 9

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
---------------	----------------	--------	----------	-------------

No matches found

Search completed: April 28, 2003, 11:02:43
Job time : 16.6667 secs

THIS PAGE BLANK (USPTO)